AD-A077 920 SYSTEM RESEARCH LTD RICHMOND (ENGLAND) F/G 5/1

CURRENT SCIENTIFIC APPROACHES TO DECISION MAKING IN COMPLEX SYS--ETC(U)

APR 76 G PASK DA-ERO-75-G-072

UNCLASSIFIED ARI-TR-76-B1 NL

1 OF 2

ADA 077920

1 OF 2

ADA
077920

LEVEL II

ARI TECHNICAL REPORT
TR-76-B1

AD A077920

Current Scientific Approaches To Decision Making In Complex Systems

Report of a Conference held at Richmond, Surrey, England, 14-15 July 1975

by
G. Pask

SYSTEM RESEARCH, LTD.
2, Richmond Hill
Richmond, Surrey, England

ARI LIAISON OFFICE, EUROPE
European Research Office, U.S. Army, London, England

APRIL 1976

DDC FILE COPY

Contract DAERO-75-G072 new

Prepared for

U.S. ARMY RESEARCH INSTITUTE
for the BEHAVIORAL and SOCIAL SCIENCES
1300 Wilson Boulevard
Arlington, Virginia 22209

79 22 5 105

The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

6

CURRENT SCIENTIFIC APPROACHES TO DECISION MAKING IN COMPLEX SYSTEMS.

9 Technical rept.,

10 G. Pask

System Research, Ltd.
2, Richmond Hill
Richmond, Surrey, England

15 DA-ERO-75-G-Ø72

18 ARI

19 TR-76-B1

Accession For	
NTIS GRA&I	X
DDC TAB	
Unannounced	
Justification	
By	
Distribution/	
Availability Codes	
Dist. A	Avail and/or special

ARI Liaison Office, Europe
European Research Office, U.S. Army, London, England

12 107

Approved By:
J. E. Uhlaner
TECHNICAL DIRECTOR

U. S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

Office, Deputy Chief of Staff for Personnel
Department of the Army
1300 Wilson Boulevard, Arlington, Virginia 22209

11 Apr[redacted]76

Contract DAERO-75-G072

This is an exploratory research project designed to meet military management requirements for research bearing on a specific management problem. A limited distribution is made--primarily to the operating agencies directly involved.

339 930

JOB

List of Participants ("ARI" stands throughout for "US Army Research Institute for the Behavioural and Social Sciences")

Dr J Baker - Supervisory Project Director, Organisations and Systems Research Laboratory, ARI Washington.

Dr D Broadbent - Dept. of Experimental Psychology, Oxford

Col Freund - Commanding Officer, European Research Office of the U S Army, London.

Dr R Hogarth - INSEAD, Fontainbleau, France.

Dr D Holmes - Chief - ARI Field Unit, Heidelberg, Germany

Dr M Kaplan - Chief, ARI, Liaison Office, European Research, London.

Dr R Keeney - IIASA, Laxenburg, Austria

Dr G Kleiter - Psychology Institute, University of Salzburg

Prof D Kahneman - Dept of Psychology, Hebrew University of Jerusalem.

Dr G Pask - Director of Research, System Research Ltd, (and Dept. Cyberentics, Brunel University; Inst. Ed. Tech. Open University).

Dr L Phillips - BIOSS, Brunel University.

Dr Carl Axel S Stael von Holstein - Stanford Research Institute, Scandinavia, Stockholm.

Dr H de Swart - Vrije Universitat, Amsterdam

Prof A Tversky - Dept Psychology, Hebrew University of Jerusalem

Prof Dr D Wendt - Institute fur Psychologie, Kiel, Germany

Dr J Zeidner - Director, Organisations and Systems Research Laboratory, and Associate Director, ARI, Washington.

CURRENT SCIENTIFIC APPROACHES TO DECISION MAKING IN COMPLEX SYSTEMS

Final Technical Report

- Index -

Current Scientific Approaches to Decision Making in Complex Systems

ABSTRACT

In July 1975, a conference was held at Richmond, Surrey, England to elicit a fair picture of the state-of-the-art in decision making in Europe; the status of decision oriented disciplines; and ongoing or contemplated lines of research. Twelve papers on the subject of human decision making were presented and discussed. These papers cannot be neatly classified since one of the main points that emmerged from the discussion was that the classical departments of decision making (for example, the distinction between a prescriptive or normative theory and a descriptive theory, which accounts for the human decision process) are not usually appropriate to the complex situations of primary interest. Some papers described work with man/machine simulations. Whilst affirming the value of such arrangements as realistic to users, practicable as models and profitable as experimental tools there is an outstanding issue of the trade off between "on line" training (what is learned in the simulation) and pretraining. Further, a special theory of such mutualistic systems is required. Other papers dealt with the odd distortions (according to the "rational" standard) of human beings the heuristics that humans adopt and the virtual impossibility of distinguishing decision making, learning and the formulation of problems. There were several recurrent themes to the conference proceedings:

(1) The otherwise confusing picture of human decision making is somewhat simplified by the existence of individually and culturally consistent cognitive styles and methods of recall. Hence, given a kit of decision tools, including computer aids as well as mathematical models, it is possible either to adapt the tools to suit the human decision maker or, vice versa, to train the decision maker to use decision tools that are chosen on grounds of efficiency or elegance.

(2) It is possible that people may be trained to respond in a particular way that yields performance data: for example, to use confidence estimation and rating facilities. However, this does not mean that they can be trained to decide.

(3) Attempts to fit simple minded models to human performance are generally unsuccessful, especially for sequential and temporally dependent choice. More predictive models are available at the price of increased sophistication and properly accounting for underlying cognitive processes. At the minimum these sophisticated tools are required to underpin decision making at the level of individual or group decision.

(4) Decision analysis is not, as sometimes supposed, a prescriptive exercise, introducing the decision maker to what he should do. Nor, as sometimes supposed, is it descriptive. The chief merit of the method, which has attestable value when used in this spirit, is to establish a compromise between the decision maker ~~(team or individual)~~ and the expert, so that, as a result of learning and dialogue, it is possible to construct a compromise model, ~~(involving structural, probabilistic and utility components)~~, which is on the one hand acceptably rational and communicable, on the other hand, tailored to the decision makers' ingrained habits. It was suggested that man/machine systems could be analysed and designed from a similar point of view.

1. Overview

1.1. Background

The conference took place over two days and three nights at Richmond. Our residence was the Richmond Gate Hotel. It used to be the Moorshead Inn; though the bedrooms are modernised, the front architecture and atmosphere are little changed from the days when Mr Tupman (the lover amongst the members of the Pickwick club) retired to this Royal Town. The deer park spreads out on our left, some 8 square miles of it, wild enough in feature to charm the countryman, but carefully landscaped, to suit Johnson himself. In front, the hill dips down abruptly to the Thames and up the river, in fair weather, you can see past Hampton Court and as far as Maidenhead. On the right, the houses and hostelries that crown the hill retain a mellow elegance that comes, to my mind, only from those 18th century code books of construction practice that turned the smallest builder into a minor architect; outstanding, pedestrian, or downright shoddy; but always with a certain style.

The participants, observers, organisers and staff half filled the hotel, so that the feeling approximated closely to a country house party. We met and dined together; Mr Dennis is an excellent butler and his ladies and gentlemen are most courteous. The occasion had neither the formality or the expansiveness of a gathering at Ditchley House, but it did have essentially the same spirit.

So much for the setting. In all, there were 12 papers and several sessions for debate and discussion. As a rule, the papers were interrupted by questions or comment and sidelines were pursued. Moreover, as should be the case, just as much emerged from exchanges outside the meeting room as those within it. Towards the end of the proceedings, I was asked to write an overview or summary and allowed the liberty to cull material from various places. Quite often my remarks refer to off-the-record interchange and they do not always respect the order of presentation. In particular, the passages labelled comment are reports on personal conversations and other peoples conversation, or even meeting room asides. They appear significant enough to include as part of the overview report, but, clearly, I may have half heard or misinterpreted what was said and I am alone responsible for the mistakes, if any, thus introduced. Other comments, chiefly in footnotes, are intended to make the text more easily readable to the non-specialist.

The entire proceedings in the meeting room were recorded and transcribed by Judy Fagelston and have been edited, for the record, by Elizabeth Pask and Linda Barsby, points of ambiguity being resolved by Dr Bernard Scott. These records are filed with Dr Zeidner and Dr Baker, in Washington. Valery Robinson checked through the summary report.

The idea of the conference was discussed in 1974, at a meeting between Dr Zeidner, Dr Kaplan, Elizabeth Pask and myself. Dr Hill gave his advice regarding a propitious organisation and Dr Zeidner, Dr Baker and I later met in Washington to firm up a plan for submission to Colonel Freund.

Quite clearly, the patrons have a special interest not only in decision making, decision formulation, and learning how to decide, but also, in man/man symbiosis (team decision) and man/machine symbiosis (computer assisted decision, either at the strategic or tactical level). Another goal, however, was to elicit a fair picture of the "state of the art" in Europe; the status of "decision oriented" disciplines; ongoing, or contemplated, lines of research and, so far as possible, the success and professional acceptance of any applications of these methods. Such objectives are likely to conflict if the entire meeting period is short and two innocent tricks were played in arranging the meeting.

(a) The man/man and man/machine aspects of decision making were not emphasised excepting in the Chairman's opening comment and the first paper by Dr Baker (b) Insofar as it was possible to tell what the participants would say (for the brief given to the invited experts was very broad) papers of all types were distributed over both days in a fairly successful attempt to induce a spiral or re-entrant structure, rather than a series of discrete and specialised discussions.

As a result of this, one participant (Dr Phillips) expostulated "If you had asked me to talk about neat ways of using computers to help in Bayes estimation I could have given you lots of them". True, and much the same might have been said by most. In a sense, however, this is precisely the kind of special orientation the structuring was designed to avoid. The aim of the conference was to bring together thoughts from diverse and carefully chosen authorities, to unify them so far as possible and to debate their mutual relevance. If and only if decision aiding programs were regarded as paramount by the participant's would it have been proper to highlight them as salient topics for debate.

In fact the discussion did turn through a full circle. Starting with Dr Zeidner's survey of the field and Dr Baker's presentation of SIMTOS (a complex, computer aided, system for simulating tactical decision situations) it passed through many areas and returned to the origin with a general recognition that, for nearly all the facets of decision which had been revealed, man/computer systems and mechanically monitored man/man systems were of outstanding interest.

(1) For research into decision formulation and choice mechanisms

(2) For operational application of the principles and theories put forward

(3) For implementing models

(4) For doing the otherwise impracticably arduous chores which attend real life applications of decision theories

(5) For adding a fresh and significant dimension to the notion of "Deciding".

This last conclusion, though expressed less volubly than the rest, was generally felt, outside the meeting room, to be the most important of all and to open up the possibility of a genuine, practically useful, and unified, theory of "decision".

1.2. Organisation of the Summary Report

To preserve individual identity the presentations are listed in proper order (Section 2) together with brief notes intended to capture the gist of their content and the discussion (in which Dr Kaplan and Dr Holmes took an active part, though they did not read papers).

Due to the spiral structure of the proceedings, categories of study are often obfuscated. They are retrievable, so far as they exist, from Fig 1 which allocates papers to fairly well acknowledged disciplines.

Sections 3, 4, 5, 6, 7, 8, are a series of essays intended to summarise some of the salient issues, and, in some cases, to point out unavoidable omissions which should provide guidelines for the organisation of any further conferences. Insofar as I am biassed (towards cognitive processes and their exteriorisation), these essays are biassed. I have done my best to express consensual ratings of significance, but the bias is present and should be kept in mind.

	Utility	Probability and confid-ence estim-ation	Decision Analysis	Man Machine systems Gaming and game sim-ulation	Team Operation
Authors and their papers ref erenced in Section 2	Kahneman Keeney Wendt	Baker De Swart Kahneman Kleiter Phillips Tversky	Keeney Stael von Holstein	Baker Broadbent Phillips Zeidner	Baker Broadbent Hogarth Zeidner

	Training and remedial measures	Experimental Studies	Cognitive Processes	Decision Styles	Mathematical Theories
Authors and their papers refer enced in Section 2	Baker Broadbent De Swart Keeney Tversky Stael von Holstein	Broadbent Keeney Kleiter Wendt Stael von Holstein	Baker Broadbent Tversky Zeidner	Baker Broadbent Kleiter Phillips Tversky	Kahneman Keeney Kleiter

Section 1. Fig 1.

2. Outline of Presentations

2.1. Dr Zeidner: "Introduction".

After dealing with procedural matters and delivering the appropriate courtesies, Dr Zeidner presented a bunch of focal interest points; speaking now, in the role of a behavioural scientist. He clearly has Elfin as well as scientific ancestry and the following dead-pan condensation does scant justice to his remarks.

There is a real and pressing problem, at all levels of command and control, of establishing effective mutualism between groups of people and between human beings and computing machines. The power of the machinery only becomes available if decision processes are rationalised (spoken of clearly, stated as heuristics). Conversely, human misperception either of machine capabilities or the decision processes set in motion, leads, just because of the power and partitioning that goes with mechanisation, to potential but hitherto unfamiliar hazards.

Under these circumstances, it is essential to heed the process of choice as well as its outcome; the cognitive or pragmatic operations which result in votes, actions, or moves. It is also important to recognise that "Decision" involves problem formulation as well as hypothesis testing and this, alone, places information processing fairly and squarely in the domain of decision making. Because of that a number of apparently peripheral questions are actually of crucial relevance; for example, data base design and the appropriate redundancy of data representations, information sources etc.

The human components of command and control systems are traditionally considered under "selection" (of "good" decision makers) and "training" (to convert ordinary decision makers into "good" ones). Surely, both of these departments are pertinent, but several factors, salient if the human beings act as part of a mechanised or team organised system, tend to modify the traditionally established departmental boundaries.

(a) It is virtually impossible to conceive of any real life decision process that does not entail learning, and various auxiliary types of problem solving.

(b) In this context, specific decision styles exist and are generally idiosyncratic (individual differences being more or less stable).

(c) There are few adequate criteria of what makes up a "good" decision maker, when choice is considered as an isolated event. The acceptable criteria of "good" are operational and bound to system performance.

Given (a), (b) and (c), it appears that training, rather than being regarded as a separate activity, may often be profitably incorporated into the system itself (a sophisticated form of the "on the job" training philosophy); so, bearing (c) in mind, may

2.2. Dr Baker: "SIMTOS - a review of Recent Developments".

SIMTOS is a command and control simulation involving both human beings and computing machinery. The hardware (computers, terminals displays and so on) occupies about half a floor of a sizeable office block. The simulated world is military; the data and action structure was developed in the context of NATO, using information provided verbally and by questionnaire responses from experienced officers (as later, this data structure was subsequently condensed and streamlined).

The "world" is simulated at the level of a tactical operations centre and based upon an actual territory (for example, place, names, distances, etc., are veridicial). The simulation is dynamic, since it responds to actions and, in addition to the "friendly" forces and resources it contains a stochastically simulated "enemy" force. * In this "world" there may either be a combat oriented or a defence oriented situation which prevails for 24 hours of real time or (on the accelerated time scale which can be employed in the laboratory, since it is possible to present information without delay and to focus upon decision making) 8 hours. In either case, a typical scenario ordains an initial period of data gathering and planning, followed by a period in which emphasis is placed upon action and reaction to an ongoing feedback regarding the results obtained.

Subjects are treated individually and emulate a field commander in the following, crucial, sense.

It is not possible to gain direct access to the decision process taking place in a general's head. However, due to the organisation in which he works, the general exteriorises much of this decision process through his staff officers, who are responsible not only as data collectors, and executives, but also for contriving solutions to problems as they appear. According to one's disposition in the matter it is plausible to say "the activities of the staff officers <u>reflect</u> what is going on in the general's mind" or that a general's mind is "<u>made</u> by his staff (and its quality reflects the quality of his staff)". Probably both statements are true, and may, in practice, be equivalent. Amongst the staff of a command centre, the operations and planning officer (G3) and the intelligence officer (G2) are pivotal and, in reality, these and other positions are occupied by different people. However, all professional officers have basic cognisance of each role. The philosophy underlying SIMTOS is that one subject can be caused to play several parts (G3, G2, for example) by contriving suitable message structures and using mechanically executed routines to perform duties complementary to his immediate role. Insofar as staff activities exteriorise a commander's decision process, and vice versa, this expedient allows the individual subject to act as a commander and records of the subject's transactions characterise a tactical decision process.

---

* This is the "one sided" form of SIMTOS, chiefly discussed. The system may also be used with two opposing teams in which case the "enemy" is one of the two.

"evaluation". Further, from point (b), high quality decision performance is likely to depend upon matching human decision makers to mechanical aids and/or to other human beings.

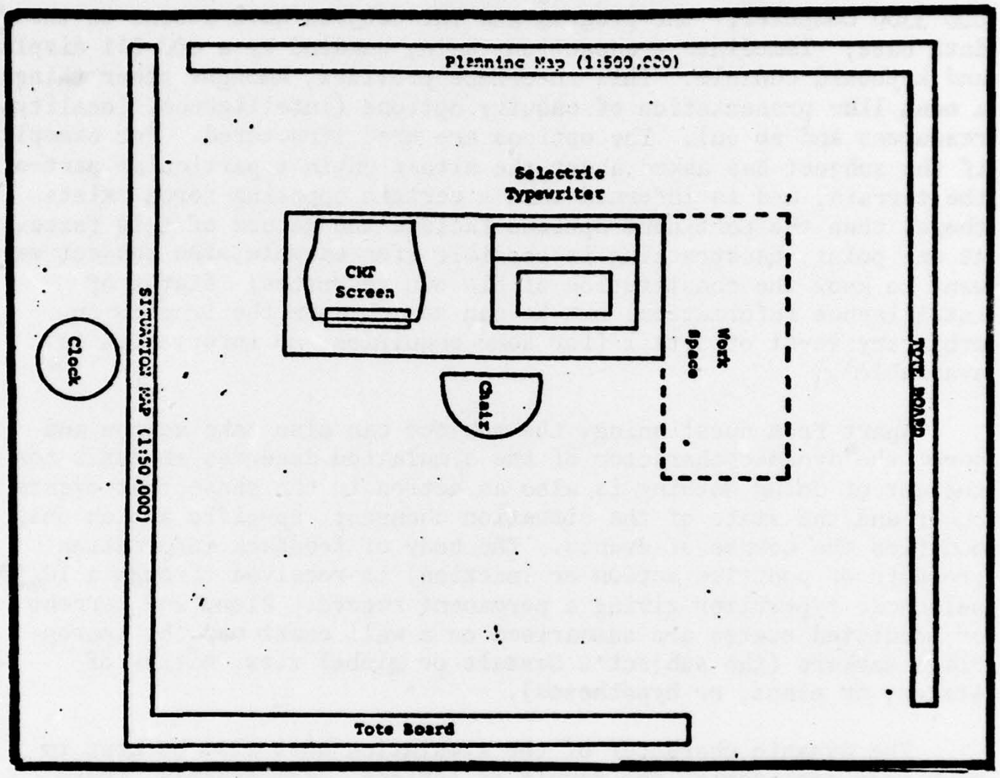
Overriding all this, certain basic limitations exist. For example, human beings are prone to "risk distortion" (over cautious, or under cautious), to ingrained misperceptions about "chance" and they have a restricted ability to recognise and recall. Machines, of course, have their own limitations but fortunately they are different ones. Provident system design seems to depend upon a subtle deployment of the matching and interactive possibilities voiced in (a), (b), (c) and recognising the constraints imposed upon human beings and machines so that the system is biased, to assign humans the tasks they <u>may be</u> "good at" doing, and machines the very different tasks they <u>are</u> "good" at doing.

After a brief engagement with the projector (from which he emerged, victorious) Dr Baker described the subject's task in some detail; Fig 1 shows a subject's operating station. The body of SIMTOS simulating a tactical operations centre, is implemented on a CDC 3300 computer; the program and the subject have access to the data base; immediate transactions being handled by a CDC 211 display and keyboard console. This interface provides, amongst other things, a menu like presentation of enquiry options (intelligence, locality, resources and so on). The options are tree structured. For example, if the subject has asked about the situation in a particular part of the terrain, and is informed that a certain opposing force exists there, then the pertinent options include the nature of this force. At any point, backtracking is possible (for example, the subject may want to know the constitution of his own resources; status of intelligence information) but he can search down the tree to an arbitrary level of detail (for some enquiries, no information is available).

Apart from questioning, the subject can also take action and here, the dynamic character of the simulation deserves emphasis for the act of doing nothing is also an action in the sense that events occur and the state of the situation changes; specific action only modifies the course of events. The body of feedback information (results of positive action or inaction) is received through a 10/50 selectric typewriter giving a permanent record. Plans and current or predicted states are summarised on a wall chart map, by appropriate markers (the subject's Gestalt or global view, either of states, or plans, or hypotheses).

The dynamic character of the simulation must also be kept in mind when considering the result of actions. For example, there are routines which move soldiers from one location to another, movement time depending upon transport and obstacles in the way. There are combat routines to determine the outcome of an engagement, based upon veridicial estimates of attrition. These, in turn, call specific subroutines; for example, casualties suffered if an action is taken given the existence of weapon configurations which may (or depending partly upon his expertise, may not) be known to the subject. The time at which messages indicating the results of action are received is based upon statistically veridicial estimates.

Since the interaction is computer monitored it is easy to obtain indices such as those shown in Fig 2. But the crucial question of scoring (operational effectiveness evaluation) has two distinct facets due to the fact that any run (is divisible (as noted earlier) into a preliminary phase during which the subject searches the data base and determines the current state of his environment (both his own resources and those deployed in opposition) and a later combat phase, in which the stochastic algorithms come into operation and the subject is bombarded by a mass of information as a result of his own action. Time is crucial and delays are exaggerated. Whereas various criteria of adequacy may be used in the initial phase, to determine sensible levels of data base search (for example, Bayesian stopping criteria, or the judgement of experienced officers) combat phase evaluation is a different matter for all those idealised principles and opinions are subordinate to the time or lag constraints; quoting verbatim.



Section 2.2. Fig 1. Schematic Illustration of Subject Station

INFORMATION PROCESSING FUNCTION	SYSTEM PERFORMANCE MEASURES		INFORMATION QUALITY MEASURES	
	THOROUGHNESS	RESPONSIVENESS	COMPLETENESS	ACCURACY
DISSEMINATION	Percentage of messages which arrive at the proper destination	Speed of message preparation, processing and routing.	Resistance to information losses during dissemination.	Correctness of disseminated information.
COMPUTATION		Speed with which computational reports are requested and prepared.	Completeness of computational reports.	Correctness of computational reports.
COMPILATION		Speed with which Summary reports are compiled.	Completeness of compiled Summary Reports.	Correctness of compiled Summary Reports.
RETRIEVAL		Speed with which queries yield responses.	Completeness of responses to queries.	Correctness of responses to queries.

Section 2.2 System measures for evaluating message processing performance and information quality

"One forgets that events occur rapidly, but the reactions occur slowly .... events happen rapidly when one is fighting against them and the actions one takes seem to run slowly so the decision maker faces a nightmare like dream which builds up around him".

In this phase performance (in terms of loss, casualties, territory acquired and so on) is scored against a standard obtained by "letting the system run without decision makers" i.e. by taking the subjects situation at the beginning of the combat phase as an initial condition and running the inbuilt offensive against the inbuilt defensive algorithms, in many combinations, stochastically.

One impressive feature of the system is that professional subjects accept it as realistic and become trapped or entrained by its operation. This alone justifies the underlying philosophy that it is possible to exteriorise and record a tactical decision process. Efforts to work out performance predictor variables have met with some (but limited) success; however, it is possible to identify specific decision styles both in the initial and the combat phases of operation. One especially interesting finding is that realism can be achieved by using a relatively impoverished data base and display. For example, when officers were asked to comment upon and criticise the wall chart maps (used to summarise states and the subject's Gestalt like overview) they typically asked for all manner of detail. As a matter of fact a cartoon representation will suffice and is all that can be handled in the combat phase. However, the feature selection (what is represented) is quite idiosyncratic. Again, unless a suitably condensed global picture is assimilated during the initial phase, performance in the combat phase is impaired by overload ultimately by "sacrificing" certain units; that is, simply "disregarding" them.

Comment: At this juncture the conference participants had established a common language and several points were debated. For example, Dr Kahneman took up the issues of information requirement problem partitioning etc mooted in the last paragraph and agreed that the findings were in accord with recent research in other fields. There was some discussion of "subjective probability" or "confidence" estimation bearing both on SIMTOS and Dr Baker's pioneering work in this area, 10 years ago. We return to these considerations in Sections 3, 4, 5, 6, 7, and 8.

The following comments are, however, essential to establishing continuity with the other papers.

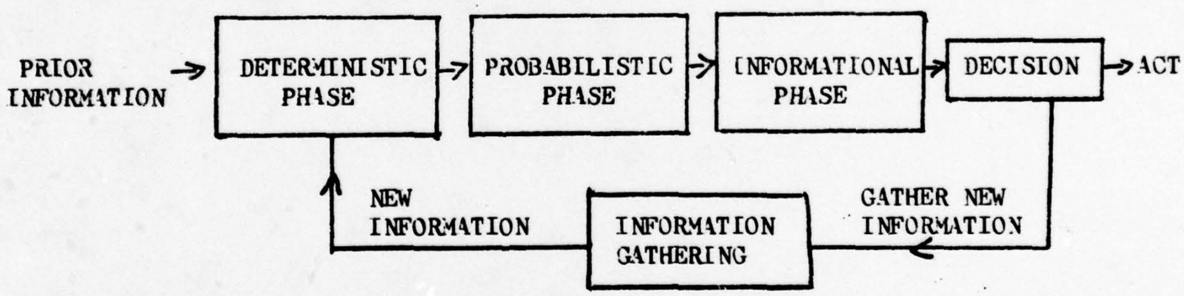
(1) SIMTOS represents a trade off between real life fidelity and laboratory control. The question of whether or not a satisfactory trade off is attainable (which had been raised by Dr Zeidner and others) can be answered in the affirmative, by noting the psychological realism of SIMTOS, as seen by diverse professional subjects, and noting that their exteriorised mental operations can be and are objectively recorded. Experience with similarly-structured non-military systems (admittedly, much less complex) was cited in support of this point.

(2) SIMTOS was adopted as the paradigm for a tactical decision situation (in contrast to the strategic decision situation typified in the next paper of Section 2.3). The tactical/strategic dichotomy is valid enough, but the intended distinction is really between a situation in which the decision maker acts on line and often under time pressure (the tactical situation) and a situation in which the decision maker has as much time as he needs at his disposal, to ruminate, or plan, or apply mathematical methods (the strategic situation). On these grounds, the initial phase of SIMTOS counts as strategic (at any rate, some subjects behave as strategists) though the latter phase is unequivocally tactical. Later in the proceedings, the participants tended to use the tactical/strategic distinction somewhat differently: to dub all situations in which cognitive processes can be exteriorised as observable behaviours, "tactical" regardless of the pace, or pressure, involved. In retrospect, this is misleading, for example, the subject's mental activities are just as much exteriorised in the initial (strategic) phase as they are in the tactical phase. By the same token, given the necessary apparatus, the strategic decision process of the next paper could also be exteriorised.

2.3. Dr Carl Axel S. Stael von Holstein "Modelling Complex Decisions and Probability encoding".

Dr Stael von Holstein is primarily concerned with models, either computer simulated or not, which embody policy or can be used to aid policy formulation in the context of large systems; for example, the introduction of new products, investment in production facilities, or the control of brush fires in California. Insofar as the strategic/tactical distinction (comment to Section 2.2) is accepted, the process of setting up a model and ultimately using it is a definitely strategic activity. The essentials of the process called Decision Analysis are shown as an idealisation, in Fig 1.

The boxes represent phases in an ongoing process which is iterated until a satisfactory model is achieved. (A) Initial data are given or collected regarding the system of interest. (B) During a deterministic phase, the analyst formulates a model structure which he believes is likely to describe the way the system works. In general, these structures (the abstract statements of the footnote) leave out unimportant or seemingly irrelevant particulars, variation in which will not affect the criterion variables if the mechanism is executed. Moreover, certain of the inputs and the initial conditions are (either in principle or practice) indeterminable.



Section 2.3. Fig 1

---

* Comment: there is an almost unavoidable confusion over the term "model". During the conference, the word was used in at least two different ways. (1) As an abstraction from concrete reality, usually a mathematical abstraction (2) In the sense of "model theory" to mean the realisation or the interpretation of certain abstract statements.

Here, usage (1) is clearly in mind. However, the two meanings ((1) and (2)) of "model" are easily reconciled. After all, a predictive model of the kind under discussion has a dynamic and, in one sense concrete, interpretation (for example, in a computer program) whenever it is used. This is a model, in the "model theoretic" sense, of certain abstract statements which the model maker has made about a different concrete interpretation, namely, the real world of transportation or business.

Issues of valuation, i.e. comparison among different criteria, are introduced in Stage (B). The deterministic phase uses sensibility analysis to determine which factors are critical to the decision at hand.

(C) The analyst tries to account for the uncertainty in the critical factors during a probabilistic phase by assigning cogent weights, likelihoods and probability distributions to these variables in the deterministic model (the hypothetical mechanism). Issues of utility may be introduced in Stage (C). The result of the analysis in Stage (C) is an evaluation of which alternative is best on the basis of the present state of information. It is often possible to gather further information before the final decision has to be made. (D) During an information phase, the now probabilistic model is executed, in order to determine the value of gathering more information for one or more variables. (E) The execution, or series of executions, is determined in a Decision Phase. Either (F) some action is taken (operating upon the real system according to the precepts of the model) or (G) further information (now more specific than the original data of (A)) is gathered and used to update the model, in (B).

Dr Stael von Holstein was unperturbed when his "Deterministic" phase was christened the "artistic" phase by several participants. He agreed that a meta-theory of hypothesis and hunch-formation should exist but does not, at the moment, do so. Further, the boundary between the deterministic and the probabilistic phases, is largely a matter of convenience. Given these caveats, he went on to stress certain desirable expedients applicable to both phases of Decision Analysis and bore the brunt of lively interpolations which, at this stage in the proceedings, had fortunately become frequent enough to give the meeting the calibre of a seminar.

The cycle of Fig 1 is often referred to as a blend of System Analysis and Decision Theory. In practice the model certainly depends, for large systems involving real decision makers, upon data obtained from experts (backed up, of course, by some knowledge of the real system to be regulated; the "initial phase" of Fig 1) and interview techniques are unequivocably superior to the use of questionnaires. This fact is not, perhaps, surprising but the greater reliability and consistency of data obtained by interviews is impressively demonstrable. Before the interviews begin it is assumed that goals i.e. to predict or to optimise or regulate, are clearly specified.

During the deterministic (alias, artistic) phase, the interviewer and the expert respondent (preferably, a real decision maker or a group of them) winkle out the structure of the model and the interviewer employs interrogation procedures involving denial and contradiction (Comment: Akin to motivation research or interrogation techniques) to ensure that sufficient factors are taken into account. Typically the structure of the deterministic skeleton is represented as a flow chart and its sufficiency is determined by enquiring whether inputs to a block are sufficient to account for the outputs. Conversely, logical demonstration is employed to remove redundancy due to gross over-specification. Finally, a crucial distinction is established between "state variables", meaning, in this context, variables over which the

decision maker has no control and "decision variables" which he is able to control, though with differing certitude and efficacity. The chance variation in the critical state variables is absorbed in the probabilistic phase. The general canon of adequacy for a deterministic model is a partial sensitivity analysis, (i.e. noting the effect of varying one, or n-tuples, of quantities upon goal associated quantities, such as profit).

The probabilities appearing in the probabilistic phase refer to the decision makers belief and uncertainty; in other words, decision analysis calls for a Bayesian (degree of belief, personal state of information) perspective, rather than a frequentist perspective (probability is the limiting value of a long run frequency). Several devices are conveniently employed to elicit degrees of belief. These are fairly standard; for example, the "probability wheel" (the respondent agrees that segment sizes, hence, the probability of a spinner-stopping in one segment, correspond to his estimate); Bayes estimation, lotteries*; hand sketched prior probability distributions and the assignment of probabilities to conditional choice trees representing possible states of the model. The values attached by the decision maker to different states or outcomes are elicited by scaling and preference rating techniques. The successful use of both these tools again depends upon the feedback available in an interview.

Since realistic models are complex, and since human beings find it difficult to estimate their doubt or certainty over numerous and highly contingent outcomes, (and, for that matter, to *determine* whether these outcomes are, as they must be analytically, sets of exclusive and exhaustive alternatives) various expedients are used to partition the model and, so far as possible, to minimise dependencies, i.e. conditionalities. For example, in considering sales output and market growth, a factory output model may be partitioned by presenting a graph of the time varying quantity (sales) with potential market growth, factoring out the growth component, re-estimating from a fresh origin and finally returning to a situation with the growth component included, but using the revised time trend estimate.

Several points of discussion emerged from this presentation. First (Kahneman, Tversky, Zeidner) there appears to be a supposition that the model is partitionable or decomposable into "Black Box" subsystems. Dr Stael von Holstein agreed that this is an hypothesis underlying most of decision analysis (and corresponds, roughly, to the lumped equivalence theorems of network analysis). Comment: the crucial feature appears to be that when decomposing a model by removing or obscuring conditionalities no dependencies liable to change the sense of probabilistic feedback, are unwittingly

---

* Let X and Y be two (in general n, for small n) events of which the subjective probability is unknown. Let the lottery $\langle X, \text{Win} \rangle$ or $\langle Y, \text{Lose} \rangle$ be presented, as a lottery A. Consider a lottery B in which it is possible to rationally vary the probabilities of events U and V (*for example, Balls from an urn, segments on* Stael von Holstein's spinning wheel) and let "win" and "lose" have the same significance in lottery A and lottery B. The subjects belief in X as against Y (say $p(X)$, $p(Y) = 1-p(X)$) is discovered by adjusting the determinable probabilities $p(U)$ and $p(V) = 1 - p(U)$ (for example, by adjusting the spinning wheel segments until the subject is indifferent between choosing lottery A or lottery B.)

eliminated in the process. A propos the same point, it appears that a theory of approximations is urgently required (<u>Comment</u>: Mostly outside the recorded proceedings, de Swart and others discussed various kinds of Fuzzy Set and Fuzzy System theory, as appropriate candidates for a theory of approximations: other participants, the closely related theory of tolerance automata). Many of the participants raised questions about the use and formal applicability of "sensitivity analysis" (one of the refinement methods) and Stael von Holstein stressed that the technique is employed iteratively (at all stages of complexity). Other issues were the extent to which experts tried to overstate the required complexity and to reduce it as they realised that a simpler representation would be adequate (a further result of interviewing strongly confirmed by several participants). Doubts were raised about the possibility of estimating time variable quantities (Baker, in particular) and Hogarth cited variations in value estimates as a function of time, due to adaptation and habituation effects.

Some of these objections are not completely answerable. But Dr Stael von Holstein pointed out that the Decision Analysis methods make difficulties common to <u>any</u> forecasting method unusually explicit; for example, the need to invoke a meta-theory of the intuitive component for the deterministic (alias, "artistic") phase is to be reduced and the explicit recognition of "Black Box" decomposability, which, depending upon the problem at hand, may or may not, be rightly assumed to hold true. In summary, apart from the manifest utility of decision analysis as a method for prescribing predictive and control strategies, use of the techniques employed for modelling bring to light a number of ubiquitous and important phenomena; amongst these are the superiority of interview, in contrast to questionnaire methods, conditions under which human beings can, and cannot, be expected to furnish reliable estimates of doubt; expectation, etc., as well as the human decision makers dependence upon data sources and predictions or extrapolations referring to a system's future state.

2.4. Dr D.E. Broadbent: Notes on Deterministic Aspects of Decision Making

Experienced managers were observed playing 3 types of business game involving multi stage decision, usually by a team, the results of decisions being fed back, periodically, by a computer. One game images a manufacturing process, another is a negotiation exercise (selling expensive installations to a developing country) and the last is control over a simulation of the British Economy (decisions are made about taxation rate, government expenditure and regulation of monetary supply; the feedback variables are Gross National Product, unemployment level, and so on). All the situations involve time of decisions leading to results and, in this sense, are "tactical" in calibre though a "strategic" element is entailed in planning.

Although these situations have unpredictable and probabilistic components, Dr Broadbent was struck by the extent to which determinate factors also influence cognition and behaviour. In general, attention is focussed on keeping track of determinate but changing values of several variables. The distinction between determinate and probabilistic factors is cited verbatim.

"What do I mean by determinate factors? For example, if someone in the first game recruits staff, there is a probabilistic element in the situation in that he does not know how many of the staff will leave when they have been trained and go off to work for another firm. That is uncertain. On the other hand, given that he knows how many people he is employing, he does know how much the salary bill will be for the next quarter. Similarly, if so much production is scheduled, he knows what will happen to stocks of materials, how much will be available to sell, and so on. These quantities are completely determinate, and yet people are having trouble keeping track of them. The data base is constantly changing. They do not carry its state in their heads, and they have difficulty following through the implications of changing states".

The following tendencies are particularly interesting (a) A clash between taking decisions intended to optimise short term performance and decisions which lead to the acquisition of information about the determined (but unknown) system and its environment. One salient feature is that optimising decisions usually change the values of several variables so that feedback is ambiguous: ideally, the quantities should be manipulated independently. (b) Difficulties over keeping track of remote contingencies, typified by budgeting in respect of a time sequence of events (payments and expenditures, for example) that reflect upon and influence an asynchronous process going on in the background (for instance, installation). (c) The relative merits of "making trials to determine the consequences of a decision" and "deciding upon a goal after which a simulation proposes means for achieving this goal". On balance the latter arrangement was recommended but an interesting counter argument was cited (due to Prof Rosenbrook, of Manchester); namely, that the mechanisation introduced by recommending actions based upon the simulated consequences of a chosen goal may lead to job impoverishment.

Various, more tightly controllable, laboratory experiments have been run with these case history observations of business

gaming in mind. For example, it appears that the doctrine of participatory learning (which underlies business gaming) does not always lead to desirable results (though the effacicity of participatory learning in some situations was not questioned). The participants do not seem, for example, to gain a concept of salient relations in the system (for example, that certain factors influence inflation) though they may acquire operational expertise. Laboratory work suggests that this may be due to the fact that the activities involved in participation (verbalisation and interruption) tend to cluster data in a manner that inhibits its memorisation.

Other experiments are concerned with miniature gaming control experiments (one of these was illustrated using Dr Phillips as a volunteer subject). Two intriguing results are as follows: (1) So far as conceptual knowledge is concerned as much can be learned by watching the game as by playing it (2) Superficially similar decision and control games differ in difficulty and also in the extent to which a rational explanation of the system is built up, depending upon the requirements for distributing attention between input data and memorised data.

The determinate factors may refer to predicted future states or to past states, retrieved from memory. The last experiment bears upon the indexing and coding of items in memory and their subsequent retrieval in order to perform a classification and decision task. The experiments are still in progress, but a number of interesting idiosyncracies and individual differences between decision makers have already been observed. Since the following remarks refer to ongoing work, they are a slightly edited version of the transcript.

"We (Dr Broadbent and his wife) started work in November, using housewives as subjects, and gave them mail order catalogue pictures of 100 potential Christmas presents to categorise under the instruction that they would later be given pen pictures of people, like, 'well-to-do young man-about-town whose interests are collecting Chinese porcelain, and water skiing'. The subject would have to find a suitable present for this person in terms of the descriptors that they had given earlier. We let the subjects look at all the possible presents, then they went back and ran through them again, classifying them.

Since there were 100 presents, the subjects could have classified by assigning seven words on a 50% basis *. Nobody did that. On the other hand, there were people who assigned one word to each object and got up towards a hundred, but, on the whole, they did not do that either. Most subjects used between 25 and 75 different words, and they never assigned more than four words to one object, and usually three (the number of things which it is easiest for someone to remember).

When the gifts were restricted to adult female presents, the number of different descriptors used by the subjects increased.

---

* That is, by binary coding $(2^7 = 128) > 100 > (2^6 = 64)$.

A Garner type information analysis of the constraint between the words, shows that the amount of information generated in these descriptors is substantially greater than the resolution between objects that they achieve. The internal constraint between the descriptors is not increased by changing the population of gifts. The constraint looks much more like a characteristic of the individual. There is a vague relationship in that the more highly educated, the larger the internal constraint between the descriptors.

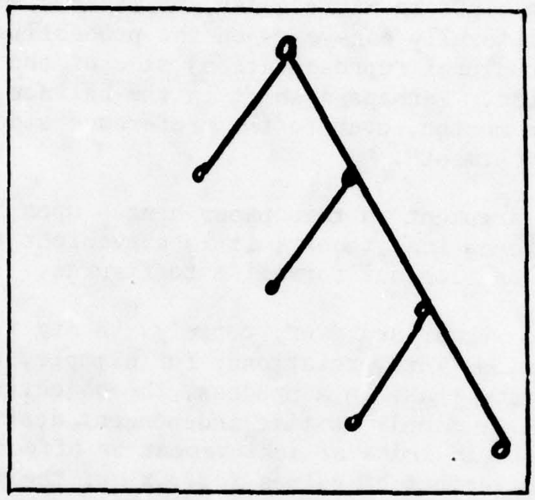
Regarding the type of constraint. (Comment: also, and because of that, the form of decision process). We have tried various possible measures, and although they work roughly, they are not very good. Intuitively, it is quite clear what the differences are. At one extreme, there is somebody who looks at a gift and says that it is so nasty that she would not dream of giving it to anyone. Of the remaining gifts, she labels some child, some adult, and sub-divides the adult gifts into male and female, and so on - a lovely hierarchical system. There were two extreme hierarchizers. One is to read Natural Sciences at Cambridge and the other is a professional librarian.

At the opposite extreme there are the people who say: this is a cheap present, or that is a dear present. This is male or female. This is practical or luxury. A complete matrix of words. Of the two people who were most inclined that way, one was a research student in history and the other was reading for a language degree. So there appear to be quite marked differences.

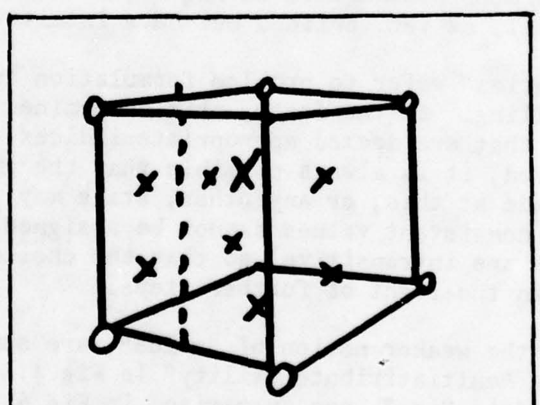
We had the same subjects back, after an interval of some weeks, and we put them to classifying thirty possible professions, and they showed the same differences in the structure of their descriptors. However, I must relate that when we tried it on a fresh group who had never done the Christmas presents, they did not do it on the professions, so it is possible that the first group remembered how they did the Christmas presents".

Dr Broadbent concluded that there are different and often consistent personal styles of data organisation, and decision making. These are believed to be "software" dispositions modified by experience but often deeply ingrained, in the adult at any rate. The extremes are parodied by Fig 1 and Fig 2. In reply to Dr Baker, recall tests showed no significant difference between the recall efficiency of different classification schemes, though, as Dr Tversky pointed out, their informational and computational efficiency is different. The stability of descriptor ordering is not so consistent as the form of the system (in response to Tversky and others). Dr Kaplan alluding both to the categorising task and the relative merits of stating goals and making trial actions, brought up the point of fitting a computer aided system to individuals or training individuals to use an "optimal" system; on balance, it seems likely that either naive or electrically minded subjects are trainable but conjectured that the habits of experienced managers are sufficiently ingrained to be immutable. Dr Hogarth, in reference chiefly to the management games, voiced the need for fitting responsibility to style, perhaps on a specialisation/generalisation continuum. Some personality/aptitude

indices do correlate with job performance at this level. But (referring to the mooted "software" nature of the internal constraints), it is an open question whether the style is invariant over a long interval. Dr Kleiter expressed interest in the hypotheses entertained by subjects in the game like laboratory situations; unfortunately, it was impracticable to elicit formal protocols, though work on these lines is in progress.



Section 2.4 Fig 1. One sided tree search neglecting choices that lead to an unwanted left note (items are sorted as "Nasty: rejected" or "nice: search progresses").



Section 2.4 Fig 2. The "matrix" type of organisation graphically represented as a multiple reference indexing scheme. Apart from graphical convenience the "matrix" is of finite but undetermined "dimension".

2.5. Dr R. Keeney: "Applications of a Theory of Decision Making with Multiple Objectives".

Dr Keeney first outlined the philosophy underlying his point of view. "As part of the motivation, there are many studies where man years are put into the modelling side of the problem, large simulations are built, many different variables are related, several scenarios run, all at relatively great expense, to produce an output. The end result is that some tables are drawn, some graphs, a large report, and often a summary report on the basis of which the decision maker or decision makers make up their minds in maybe a day, or maybe a week. The balance of effort might be literally man-years on the probabilistic and relationship (Comment : Structural representation) side of the problem to a week on the preference side. Perhaps a shift in the balance, maybe giving even a couple of man months, over to the preference side of the problem, may be worth the investment".

Comment : The argument in this paper hinges upon the practical application of a mathematical theory, it is convenient to relegate most of the mathematical and logical formalism to figures.

The notation is summarised, very tersely, in Fig 1. The objectives (O) may be given various interpretations; for example, the objectives current at a given time phase in a process, the objectives of a social, or political, group, or simply as time independent desirable goals. For any objective there is an index of achievement or effectiveness X which has values x and the vectors of values $x_1,..,x_n$ of the indices or attributes $X_1,..,X_n$ are useful for describing the consequences of a decision process.

The general approach to multiattribute decision making consists in an iteration of the steps summarised in Fig 2 (the terms appearing in this list are not all, as yet, defined but have intuitive meaning).

The "preliminaries" refer to problem formulation by discussion interview, or modelling. At the least, this determines the objectives and the attributes that are deemed appropriate indices. Since the procedure is iterated, it is always possible that the intuitive or informed choices made at this, or any other, stage may prove untenable (for example, that consistent values cannot be assigned to an attribute or that preferences are intransitive) so that the choices in question may always be revised in the light of further steps.

"Utility" and the weaker notion of "value", are summarised in Fig 3 and the notion of a "multiattribute utility" in Fig 4. The concepts of independence invoked in Fig 2, are summarised in Fig 5. The overall form of the utility function u, is restricted (as proposed, in Fig 2) by using independence principles (Fig 5) and results such as those shown in Fig 6. (Dr Keeney finds the form of Fig 6(c) especially useful).

NOTATION

OBJECTIVES		$O_1,\ O_2, \ldots . O_N$
MEASURE OF EFFECTIVENESS (ATTRIBUTE) For each of $O_1, O_2 \ldots, O_N$		$X_1,\ X_2, \ldots, X_N$
LEVEL OF ATTRIBUTE (Values of $X_1, X_2 \ldots X_N$)		$x_1, x_2, \ldots . x_N$
CONSEQUENCE	$\underline{x} \equiv$	$\langle x_1, x_2, \ldots . x_N \rangle$
COMPLEMENT	$\bar{x}_i \equiv$	$\langle x_1, \ldots, x_{i-1}, xi+1, \ldots x_N \rangle$

Section 2.5. Fig 1.

ASSESSING MULTIATTRIBUTE UTILITY FUNCTIONS

five interdependent steps

preliminaries .... terminology, aims, etc

verify independence assumptions get functional form of u.

assess single - attribute, utility functions obtain the u_i's

assess the scaling constants obtain the k's

consistency checks involves iteration

Section 2.5. Fig 2.

NEED a multiattribute utility function u over attributes X. Further ideally we want $u(x_1, \ldots, x_N)$ to be valid in the Von Neumann - Morgenstern sense so that the expected utility is an appropriate criterion to guide our decisions.

PROPERTIES OF u

1. $u(\underline{x}) \geq u(\underline{x}^x) \Longleftrightarrow x \succsim \underline{x}^x$

2. alternative A_i, i = 1, 2, has expected utility $E_i\,[u]$, then

$$E_1\,[u] \geq E_2\,[u] \Longleftrightarrow A_1 \succsim A_2$$

(Comment: "$\Longleftrightarrow$" stands for the biconditional implication "$\equiv$")

DEFINITION Any function $v(\underline{x})$ satisfying property 1 is called a value function.

A value function can be useful in obtaining a utility function

Section 2.5. Fig 3.

APPROACH Find an f such that

$$u(x_1, x_2, \ldots, x_N) = f\left[u_1(x_1),\ u_2(x_2), \ldots,\ u_N(x_N)\right],$$

where u_i, $i = 1, 2, \ldots, N$, are single attribute utility functions.

We may use an intermediary step, to find g, such that

$$v(x_1, x_2, \ldots, x_N) = g\left[v_1(x_1), v_2(x_2), \ldots, v_N(x_N)\right].$$

where

v_i, $i = 1, 2, \ldots, N$ are value functions.

EXAMPLE (Due to Peter C. Fishburn)

The utility function u is additive if there are constants k_i

such that $u(x_1, x_2, \ldots, x_N) = \sum_{i=1}^{i=N} k_i \cdot u_i(x_i)$

Section 2.5. Fig 4.

THREE INDEPENDENCE CONCEPTS

Partition $\left\{X_1, X_2, \ldots, X_N\right\}$ into $\left\{Y, Z\right\}$

so $u(\underline{x}) \equiv u(\underline{y}, \underline{z})$

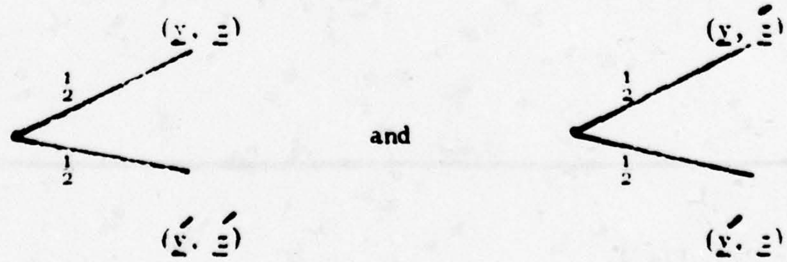
PREFERENTIAL INDEPENDENCE. Y is "preferentially independent" of Z if preferences for y - levels do not depend on a fixed level of z.

$(\underline{y}, \underline{z}^0) \succcurlyeq (\underline{y}', \underline{z}^0) \Longrightarrow (\underline{y}, \underline{z}) \succcurlyeq (\underline{y}', \underline{z})$, For all $\underline{z}$

UTILITY INDEPENDENCE Y is "utility independent" of Z if preferences over lotteries on y - levels do not depend on a fixed level of z.

$u(\underline{y}, \underline{z}) \equiv$ for all $\underline{z}$, $h(\underline{z}) > 0$: $g(\underline{z}) + h(\underline{z})\, u(\underline{y}, \underline{z}^0)$

ADDITIVE INDEPENDENCE Y and Z are additive independent, if lotteries



are indifferent for all $\underline{y}, \underline{z}, \underline{y}', \underline{z}'$

Section 2.5 Fig 5

RESULTS (two attributes)

- If Y and Z are additive independent, $u(\underline{y},\underline{z}) = k_Y\, u_Y\,(\underline{y}) + k_Z\, u_Z\,(\underline{z})$, where $u_Y\, u_Z$, are scaled 0 to 1 and k_Y and k_Z are scaling factors.
- If Y and Z are each utility independent of each other,

  $u(y, z) = k_Y, u_Y.(\underline{y}) + k_Z, u_Z(\underline{z}) + k_{YZ}. u_Y(y)\, u_z\,(\underline{z})$

  where u, u_Y , and u_z are scaled 0 to 1, and K_Y,k_Z, and k_{YZ} are scaling factors.

RESULTS (in general)

(A) Let P. I. denote "Preferential Independence".

If $\left\{X_i, X_j\right\}$ P.I. $\bar{X}_{ij}$, for all $i \neq j$, then $v(\underline{x}) = \sum_{i=1}^{i=N} v_i(x_i)$

If (1) $\left\{X_i, X_j\right\}$ P.I. $\bar{X}_{ij}$

(2) $\left\{X_i, X_k\right\}$ P.I. $\bar{X}_{ik}$

then (3) $\left\{X_j, X_k\right\}$ P.I. $\bar{X}_{jk}$

(B) If X_i and $\bar{X}_i$, i = 1, 2,..., N are additive independent then

$$u(x) = \sum_{i=1}^{i=N} k_i.u_i(x_i)$$

where u and u_i, i = 1, ... , N are scaled 0 to 1 and k_i, i = 1, ..., N are scaling factors

Section 2.5. Fig 6 (b)

(C) Let U. I. denote "Utility Independent".

If X_i is U.I. of $\bar{X}_i$, i = 1, z, ..., N, then

$$u(\underline{x}) = \sum_{i=1}^{N} k_i.\, u_i(x_i) + \sum_{\substack{i=1 \\ j>i}}^{N} k_{ij}\, u_i(x_i)u_j(x_j) + \ldots + k_{1\ldots N}u_1(x_1)\ldots u_N(x_N)$$

where u and u_i, i = 1 ,...,N are scaled 0 to 1 and the k's are scaling factors.

This form of utility function is called a multilinear utility function.

Section 2.5 Fig 6 (Contd over)

RESULT for $N \geqslant 3$, if X_1 is utility independent of $\{X_2, \ldots, X_N\}$ and if $\{X_1, X_i\}$, $i = 2, 3, \ldots, N$, is preferentially independent of $[X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_N]$,

then either

$$u(x_1, x_2, \ldots, x_N) = \sum_{i=1}^{N} k_i . u_i (x_i) \qquad \text{(Additive)}$$

or

$$1 + k\, u(\underline{x}) = \prod_{i=1}^{N} [1 + k.k_u . u_i (x_i)] \qquad \text{(Multiplicative)}$$

where

(1) $u(x_1^o, x_2^o, \ldots, x_N^o) = 0$, $u(x_1^*, x_2^*, \ldots, x_N^*) = 1$,

(2) $u(x_i^o) = 0$, $u_i(x_i^*) = 1$, for all i,

(3) $0 < k_i < 1$, for all i,

(4) u is additive if and only if $\sum k_i = 1$, and u is multiplicative if and only if $\sum k_i \neq 1$.

Then $1 + k = \prod_{i=1}^{N} (1 + k.k_i)$, to find k, where $k > -1$.

Section 2.5. Fig 6

The next step is to assess the utility (u_i ; $i = 1, \ldots, N$) of each of the attributes separately and, after that, to assess the scaling constants in the resulting mathematical expressions. *

Finally (but recall the caveat about iteration) the consistency of the normative structure erected by these expedients is tested and is either sufficiently supported or not (in the latter case the structure is revised). The method employed for this purpose is closely akin to criminal interrogation. The respondent has been persuaded by one means or another to state preferences within a system of which he is (ideally) aware. There are certain consequences, once that utilities and multiattribute utilities have been assigned, namely, that if the same respondent is presented with further pairs of alternative options he should prefer one over the other, as a result of agreeing that his previously expressed preferences are to be embodied in a formal system. Failing this a contradiction is detected. The mathematical ramifications of Bayes Rule are used to achieve much the same kind of rational framework in respect of probability estimation and the breakdown of uncertain situations (See, for example, Phillips, L.D, Bayesian Statistics for Social Scientists, Nelson, 1974).

The strength of this approach depends crucially (apart from the elegant theoretical expressions) upon two features emphasised in the talk. The first of the two is empirical; namely, the possibility of eliciting preferences at all from real respondents. Since there is ample evidence that this can be done, given circumstances conducive to rating and the estimation of belief and the Bayesian doctrine of breaking a problem down (under these special conditions of measurement) into a "probabilistic" part, and a part concerned with preference, and "value". Dr Keeney concentrated upon the other key issue, which is pragmatic. He cited a plethora of practical and very different situations which may be adjudicated in just this manner. The following extracts indicate the pragmatic flavour of the enterprise.

---

* Little was said about how these assessments are carried out but it may be remarked that the canonical method relies upon finding indifference points between the amount of an alternative-states attribute and a lottery offering constant (and known) gain or loss (over win lose alternatives) at a certain (known) probability win/probability lose.

So, for example, the decision maker may be indifferent between an amount c_1 of X_i and a lottery ticket admitting him to the gample $p = 0.4$ of win and $p = 0.6$ of losing; between an amount c_2 of X_i and a lottery ticket admitting him to the gamble $p = 0.2$ of winning and $p = 0.8$ of losing; thus in aggregate, specifying points of the u_i distribution. It is also worth remarking that less sophisticated and tedious rating and ranking techniques (one is proposed in Section 2.6, by Dr Wendt) are perfectly defensible if referred back, in principle, to basic metrical paradigms of indifference/equivalence.

By the same token the next procedure of estimating the assessment of scaling coefficients is referred to an idealised trade off between quasi independent utility values u_i, u_j and, once again, approximations are legitimate. The interested reader will find Raiffa, H, Decision Analysis, Addison Wesley, 1968, of great practical value, in this respect.

"The first case where utility functions were assessed, in the sense I am talking about, was by D. Yntema. He was concerned with the safety of landing aircraft in various situations which differed in terms of ceilings, visibility, and the amount of fuel that would remain on landing, if the landing was carried out in the normal manner. The decision makers for this problem were twenty U.S. Air Force pilots. A multilinear utility function was assumed, (Fig 6). Having obtained single utility functions (by preference over options), the same set of twenty pilots were presented with forty different pairs of situations to choose the best of the two pairs, and their selections were compared, for consistency and the model was a pretty good model for the situation".

Another example concerns the development of Mexico City Airport, up to year 2000:

"The first model dealt with engineering efficiency indices, such as access time to the airport, cost, the number of people to be moved because of airport development, etc. In the process, we learnt something about the problem and produced a model for including the attributes important for a strategic decision, such as flexibility."

The sense of changing representations as a result of dialogue with the decision makers is captured by the example of an individual dental student who was completing his term in the army, and had various options. After a good deal of background reading about his basic goals in life (health, family, peer esteem, personal esteem, total income) the respondent determined six different professions for evaluation. Remaining in the army as a dentist; leaving the army and becoming a private dentist; two fields of speciality within dentistry outside the army; management consultancy, and financial analyst. "The decision problem reduced to a simple tree with six alternatives. He assessed a utility function he thought reasonable and on the scale he was using, three of the alternatives had a utility of around 150 or 160, and the other three between 11) and 115. The difference was quite large, which surprised him. The three lowest rated professions were staying in the army as a dentist, management consultant and financial analyst. He looked at it" (Comment: this is the crucial point perhaps) "and said, he could now see why this was the case, and he resigned from the army. In addition, he was particularly surprised at the way staying a normal dentist had been grouped along with the two different specialities".

Similar comments apply to corporate problems, for example, a large firm of Californian environmental consultants and their objectives as a firm. The exercise was explicitly designed to help members of the board discuss and articulate their problems, not to choose between well modelled alternatives or plans.

"They started out with a long-range planning committee in 1972 which set up a hierarchy of objectives. The overall statement of purpose was to provide a climate whereby employers could achieve their personal, professional, and financial goals. Financial matters (like retainers, salary increases above minimum, an incentive and compensation scheme, a retirement plan) were significant variables. Other variables concerned coverage of the areas in the U.S.A. where work existed. Other variables indexed the proportion of total fees put back into doing things that were fun, (called professional development and gaining relevant experience)".

In discussion since 1972, the measures deemed relevant changed. Some variables were redundant and others ambiguous. For example, there was a subjective index of places outside the U.S.A. where work was being developed. On resolving the ambiguity, it turned out that the indicator was partly a back-up for recession in the U.S.A.

Assessments over the variables also changed during the dialogue: for example, the value of retained earnings. Much of the variation was due to people thinking in terms of quite different time trends. "If someone was thinking about one year, retained earnings could have a weight of about 0.67 given a scale of 0 to 1, whereas for a five year time period it fell to under 0.5. The reason was again pretty obvious. Over a one year period, they could make a lot of money with the thought that if in the next year, there was money over it could go into having the fun. But they were not willing to wait five years for such a trend".

In 1974 there was a major reorganisation. It appears that the dialogue based on utility assessment helped the board members and had a genuinely favourable influence upon the decision made in changing the company structure (the study is documented in a report, in press).

Much the same is true of even larger scale problems like nuclear facility siting in the State of Washington. However, value is a very tricky notion and although the problem is generally simplified and the number of variables reduced by discussion, certain hazards are present and must be avoided in the simplification process.

"Intuitively one might think (in the context of salmon fishing) that population density depends upon the number of salmon that die. But actually, each creek contains a different type of salmon, and they return to that same creek. I did not know that their genealogies were so specific. If a thousand salmon get wiped out from a stream with a thousand salmon, then something has been lost which will never come back. If a thousand salmon are lost from the Columbia River, which is where they all gather then (given that the river takes from all the streams) a particular subspecies will not be lost".

In conclusion Dr Keeney illustrated the main interpretations given to the "objectives" of Fig 1. The "time phase" interpretation is exemplified by a study of personal investment and consumption throughout life, due to Professor Richard Meyer: respondents consider their preferences for consumption and saving at various time intervals. Studies carried out with respondents belonging to different age groups indicate, for example, that graduates under-consume for a decade after leaving the university.

The other interpretations of "objective" are illustrated by work in hand at IIASA. One international study of "Integrated Energy Policy" compares the energy production goals in Wisconsin, the German Democratic Republic and the Rhone Valley, using attributes such as pollution (various kinds), energy produced, health effects, and environmental effects of energy production. In another study of salmon management in the Skeena River, the "objectives" consist in the goals or mores of interested factions and groups of people influenced by this activity; Indians fishing for food upstream, sports fishermen lower down the river (an important group because of tourism); lure fishermen and net fishermen (both commercial, working downstream) and the canning industry. The decisions to be made on the basis of utilities expressed by these groups,

concern the distribution of spawning grounds over the length of the river. The policy adopted will influence all of these groups and it is important to notice that the group activities interact both because of the geographical constraints (upstream, downstream, on the same river) and by couplings of an ethnic and monetary kind.

Several aspects of the paper were discussed, mostly during its presentation. Stael von Holstein, quite early in the proceedings, contended that the main use of a mutiattribute utility study probably lay in the discussion and systematic iteration of the steps in Fig 2. This exteriorises and possibly rationalises a decision maker's thoughts. Keeney agreed with this view; the process does "straighten their minds" and, as Stael von Holstein insisted, there is no such thing as a "best decision". The fiction of optimality is used to incite people into discussion and possibly to reach agreement, but the techniques themselves are not really directed to the pursuit of an optimality which is probably chimerical.

Further, as Baker pointed out, planners and forecasters are amongst the more important potential users, and it is in the context future possibilities that the complications are most prominent. One reaction is to opt out of multiattribute utility theory and resort to some less adequate but simple minded means of aggregating preferences and making decisions.

"Philips "could think of practical examples in the business community where the alternative to multiattribute utility analysis is cost-benefit analysis; that is, an analysis which assumes a linear model without paying any attention to the scaling problem. The Roskill Commission, set up to look at the question of London's third airport is a good example. Dimensions were assessed, values on dimensions assessed, the values were just added up as though every dimension was as important as every other dimension, as if the units of measurement and the zero point were all equivalent. That alternative is apparently far more popular in the business community than multiattribute utility analysis".

Zeidner "asked the executive secretary of one National Decision Council last December, if they used multiattribute utility in anything. She said she had never heard of it".

"I bet she had heard of cost-benefit analysis!"

At this juncture Stael von Holstein asked Keeney about the price paid for simplicity. "For your problems, how <u>wrong</u> would you be if you had assumed linear trade offs, and then, if necessary, a sort of utility function on the aggregate (transforming everything to dollars or lives or whatever) rather than measuring preference on individual attributes".

Keeney replied "I really do not know how <u>far</u> we should be wrong. If all the reasonably good strategies, are bunched together in the outcome space, then I do not think that we would be far off. But even if that was true, there is educational value in asking more specific questions, getting specific utility functions, and so on. It sensitizes the decision maker and makes sure that the answers obtained are consistent".

The process is educational (it exteriorises cognition and calls for rationality) but(Stael von Holstein) "The toughest thing is not assessing trade offs. Once people start to think about that they can do it. It is trying to get the utility curve, which involves very strange questions. You have to repeat the same procedure a number of times, and every time is equally tough".

After some detailed discussion it turned out that Stael von Holstein in his "Decision Analysis" plays much the same valuation tricks as Keeney, however, he often reverses the order. Apart from the different emphasis upon structure and preference, there is also a "general to specific" versus "specific to general" difference.

Tversky questioned, first, "The justification of introducing lotteries for the purpose of scaling". To do so may bias perception of the alternatives (especially when the options are not essentially riskless)* and, said secondly, "Your initial point was that we may be spending too much time on modelling in the probability side, and not enough on the decision making. The real cost is our time, and the research time in structuring the problem and the time spent observing decision makers and guiding them so that people do better than they would if left to their own devices. We might well adopt (as Broadbent did) a slightly different standpoint and ask where people go wrong, naturally, when they are left alone. In this case we should spend modelling time and our time on those aspects of decision where people are likely to go wrong, and leave them alone where they do a reasonably good job".

One point of debate was unique to this session and was brought up by Kahneman. In some of the examples, particularly in a political context, Who is the expert?" Political figures are very reluctant to state trade-offs between unemployment, safety and so on, in public at any rate. Who would place a price on such things? I have been told, for example, by planners, in France, that political figures will not give explicit assessments because these can always be used against them. You have to work a fair way down the hierarchy, below the political decision makers, obtaining utilities from people, who are removed from political pressures; which could render the utility assessments irrelevant. In effect we set up a dummy "objective" decision maker or utility assessor whose "values" are "independent and objective" and he is not prone to political pressures. In front of this "objective" creature there is an agent within the political system to face up to the corrupt politicians, and what is likely to happen next quarter before an election".

In principle, Keeney agreed about the difficulty. He, and Howard Raiffa (1972), wrote a paper, in this spirit on the use of formal decision analysis in the public domain.

Comment: Although it is not much discussed in the sequal the problem of "who has the preferences", remains open. An even more puzzling question is "who makes the decisions" or "where is the decision maker" which could be answered in several ways; by recourse to Baker's "General and Staff" interaction, for example, or by appeal to an organisation or a man machine system as the real decision maker. Looked at in this light, the status of preference is determined by the particular circumstances. The preferences may be the preferences of the decision maker, supposing he could be located; the preferences of people influenced by the decision (used, somehow, in the decision process) or the decision maker's estimates of or beliefs about the preferences of people who are influenced. It all depends, but the tricky point is that the "circumstances" which, if they are specified, disambiguate the kinds of preference involved, also lie at the root of the concept of "preference" itself.

---

* Outside the meeting Keeney, defended the scaling method on roughly the following grounds. The method would be justifiable given certain idealisations. To indoctrinate decision makers with these idealisations is one educational intention of the dialogue. Insofar as the dialogue succeeds, the lottery measuring stick is a tool on a par with the mathematical tools.

2.6. Dr D. Wendt: An attempt to implement multi-attribute utility theory in the evaluation of geographical areas for recreation.

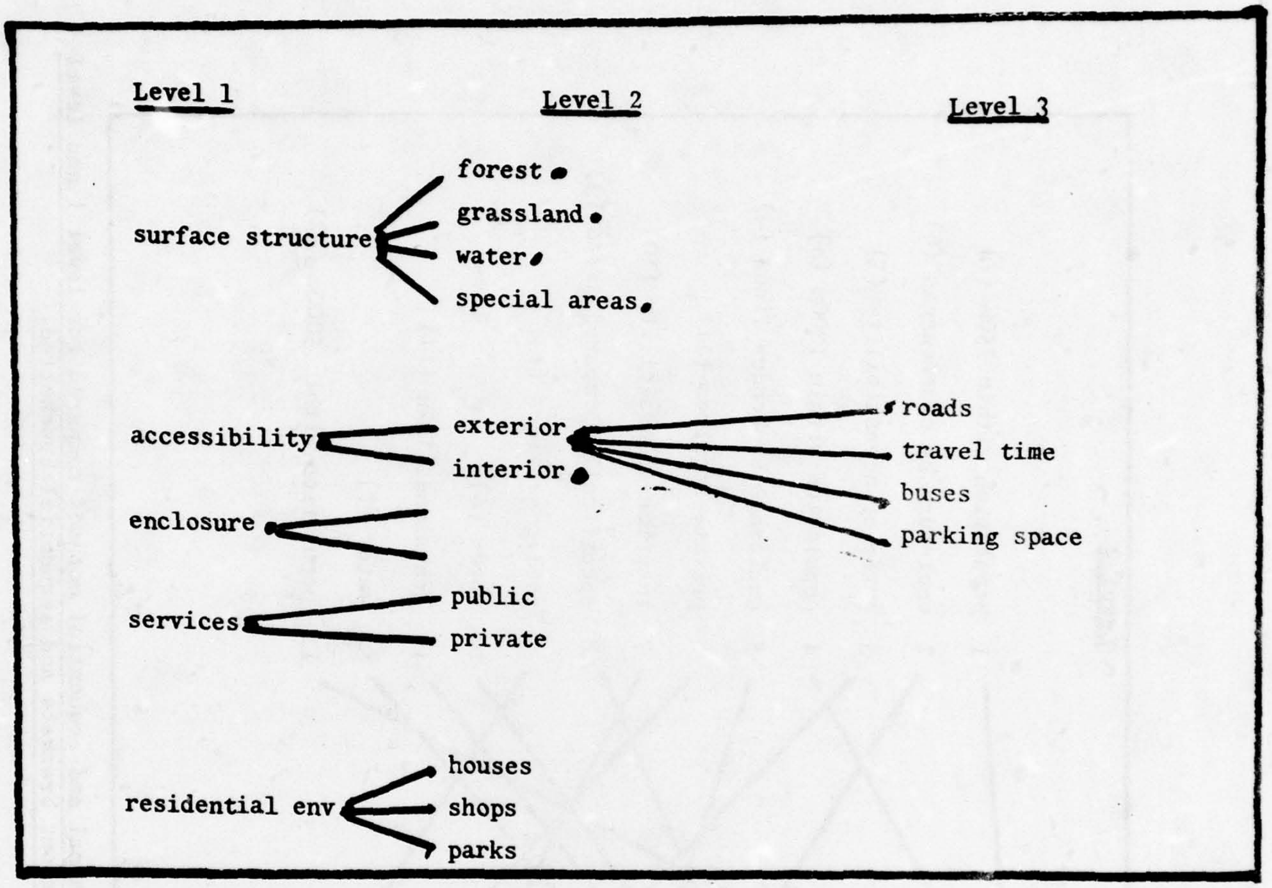
Dr Wendt described a case history in the practical application of multi-attribute utility theory. His subjects were City Planners familiar with urban development and the peculiarities of their neighbourhood. They wished to have a tool for assigning realistic values to space; to be used in planning further parks and recreational areas (the value of a region depending, for example, upon the number of inhabitants, the proximity of heavy industry and such complex matters as the connectivity of the space in question i.e. is it traversed by a transport pathway? Is it connected or connectable to a residential neighbourhood?) The generality of the space evaluation problem was emphasised; it is largely incidental that the experts were concerned with recreational usage. On the other hand, it is important to notice that they are genuine experts in the subject matter of interest.

As a first step the names of descriptive variables (specific attributes or properties) were elicited by discussion, and the sets designated by these properties arranged as part of a taxonomic classification (Fig 1). Such an arrangement is only possible given the expertise required to give proper interpretation to the attribute or property names so that the "levels" correspond to noun like clusterings. Next two methods were employed to assign "importance" weights to the properties and thus to specify single attribute utilities.

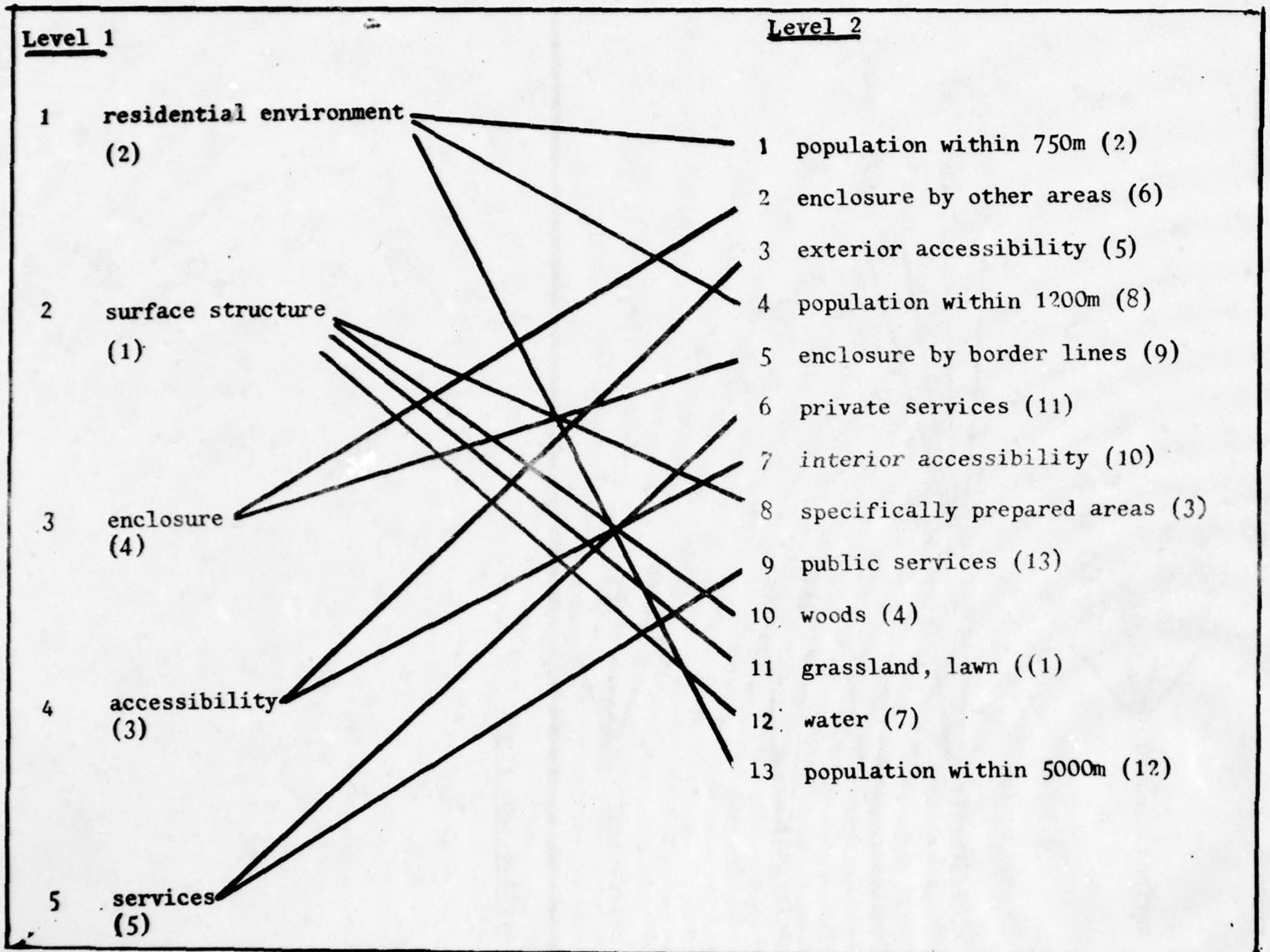
One method consisted in several (expert) judges distributing points, out of a total of 1000, over the several attributes (this is called the hierarchical method). The other method consisted in several experts rating the attribute importance on scales. Some comparative results for the two methods (hierarchical-method and rating-method) are shown in Fig 2; numerical summaries, in terms of correlation between the values ratings assigned by each method, are shown in Fig 3 and coefficients of concordance in Fig 4; in each case the data is summarised by structural level.

Using the dependencies induced by the structural classification it was possible to obtain attribute utilities for some of the attributes (Fig 5 and Fig 6). Each curve represents a utility function which is compatible with the opinion of one expert or judge, and it is evident that some plots (as in Fig 5, for example) are of the same form whereas others, (Fig 6 is an extreme case) are less consistent in this respect.

As a general comment, Dr Wendt agreed that the several experts and their peers agreed that the method gave them a great deal of insight into their own means of evaluating space: further, they felt it to be a useful device for this and other aspects of urban planning well worth the trouble of development. Formally, Fig 3 and Fig 4 show that the assignment methods are in good agreement at the highest and the lowest levels in the structural classification though not at the intermediary level. This defect might be remedied by a systemic rather than a taxonomic scheme and the disparities, evidenced in Fig 6 but not in Fig 5, can probably be traced to structural origins.



Section 2.6. Fig 1



Section 2.6. Fig 2: Results for hierarchical and sequential methods compared for level 1 and level 2 . Hierarchical without brackets and sequential bracketted.

Average of rank correlations between methods at different levels.

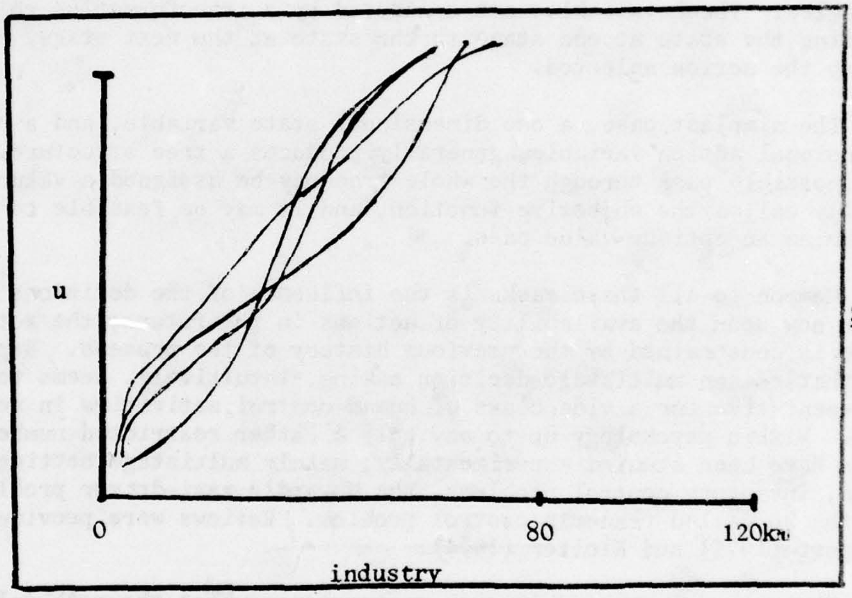
at level 1	.60
at level 2	.20
at level 3	.52

Section 2.6. Fig 3

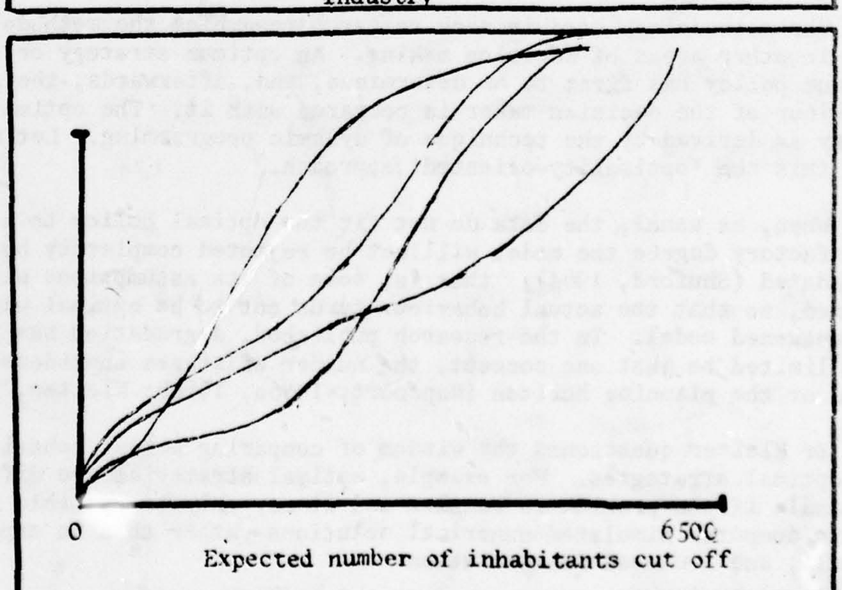
Coefficient of concordance between judgements of subjects for hierarchical(left)and sequential(right)methods.

level 1	.702	.795
level 2	.857	.209
level 3	.914	.860

Section 2.6 Fig 4



Section 2.6 Fig 5



Section 2.6 Fig 6

2.7. Dr G. Kleiter. Multiple Stage Decision Making

The Restricted Utility of Optimal Models in Multistage Decision Research

Introduction

Dr Kleiter first pointed out that the multi stage decision tasks to be considered belong to the same category as those discussed by Dr Baker (as in SIMTOS, Section 2.2) though involving less complexity. These tasks also yield information bearing upon Dr Broadbent's discussion.

In a multi stage decision problem there are a number of stages at which choice is made, a set of state variables and a set of action variables, and usually, although not necessarily, a set of random variables. These variables are connected by a transformation rule relating the state at one stage to the state at the next stage, according to the action selected.

The simplest case, a one dimensional state variable, and a one dimensional action variable, generally produces a tree structure. Each possible pass through the whole tree may be assigned a value, usually called the objective function, and it may be feasible to determine an optimum-value pass.

Common to all these tasks is the influence of the decisions made up to now upon the availability of actions in the future; the action space is constrained by the previous history of the process. Especially for that reason multistage decision making, intuitively, seems to be representative for a wide class of human control activities in real life. Within psychology up to now only a rather restricted number of tasks have been studied experimentally, mainly multistage betting games, inventory control problems, the Howard's taxi-driver problem, and the so called reader's control problem. Reviews were provided by Rapoport (1975) and Kleiter (1974).

The methodology used in such research resembles the methodology used in other areas of decision making. An optimum strategy or optimum policy has first to be determined, and, afterwards, the actual behaviour of the decision maker is compared with it. The optimal policy is derived by the technique of dynamic programming. Let me call this the 'optimality-oriented' approach.

When, as usual, the data do not fit the optimal policy to a satisfactory degree the model will not be rejected completely but degradated (Shuford, 1964); that is, some of its assumptions will be relaxed, so that the actual behaviour turns out to be optimal within the weakened model. In the research published, degradation has primarily been limited to just one concept, the number of stages considered ahead or the planning horizon (Rapoport, 1966a, 1966b; Kleiter, 1975).

Dr Kleiter questioned the wisdom of comparing actual behaviour and optimal strategies. For example, optimal strategies are difficult to handle if the problem is complex and it may only be possible to obtain computer simulated numerical solutions rather than an explicit formula, and a closed form solution.

"The behaviour of one sample of subjects is investigated in one task and compared to one optimal model or to a subfamily of suboptimal models. The deductive "one model - one task - one sample" approach is

an insufficient basis for psychological research. Investigations of this type are open to criticism (i) because they fail to provide items by which violations of the minimum standards of rationality may be checked; under the optimal model these items often fall outside the range of predictions, and thus are not included. (ii) nearly all studies neglect the reliability of the data, in respect to within-task and between-task variations; they do not vary experimental conditions by which group differences or individual differences could be related to a theoretical background".

These deficiencies are illustrated by the results of an experiment in which it was possible to establish the recommended controls and to realise the comparison of tasks and groups of such tasks. Two of the best known multistage decision tasks - the Reader's Control Problem (Rapoport, 1966a, 1966b, 1967; Toda, 196?) and a multistage betting game (Rapoport and Jones, 1970; Rapoport, 1975; Kleiter, 1975; Kleiter and Wimmer, 1975) - were investigated in a 2 x 3 x 2 design with N = 72; the three factors were: neuroticism (from a pool of 213 subjects, the 72 most extreme cases were selected), a double blind drug condition (tranquillizer 3 mg and 9 mg of a benzodiazepin-derivate produced by Hoffmann La-Roche versus placebo) and sex. While the comparison of optimal to actual behaviour seems to make sense for each isolated task, it does not do so for the overall picture of the data. Crude violations of rationality were observed, eg. choosing unfair bets or miscontrolling a system which had already reached an ideal state. Both the within task and the between task reliabilities were found to be low eg. there were no significant correlations for the efficiency over different tasks. Finally, nearly no significant group differences were found.

(a) The Control Problem multi stage decision task.

"The problem was presented in a cover-story in which the task simulates the repair of a defective machine producing pieces of a product. The proportion of defective pieces will increase over trials unless the machine is serviced. The cost of servicing which the subject may carry out at a given trial is given a definite value. Indexing trials $i = 1, 2, \ldots$, the defective pieces at trial i as X_i, and the repairing action at trial i as Y_i we have, with terms A for defect rate and B for cost of servicing.

$$X_{i+1} = \underline{A} \cdot X_i - \underline{B} \cdot Y_1$$

an appropriate loss function is (at the nth trial)

$$I = \sum_{i=1}^{i=n} (X_i^2 + Y_i^2) + X_n^2$$

The following three problems were used: Problem 1: N (number of stages) = 7, X_1 (initial state) = 2, A = 1.4, B = 0.1; problem 2: N = 7, X_1 = 4, A = 1.5, B = 0.2; problem 3: N = 7, X_1 = 5, A = 1.3, B = 0.4.

The experiments are computer controlled. Under all conditions the subject, who is required to minimise his loss, receives a tube display of the cumulative loss function, the state variable (X_i) and the action variable (Y_i) under his control. Under "practice" conditions, but not in the main experimental session, the subject also receives feedback indicating an optimal policy decision for comparison after his own decision at the trial in question. Real money is paid to the subjects according to their performance.

Analysis of variance, using data from efficiency (the ratio of actual strategy to ideal strategy), an index of over-control/under control, and response latency, showed no significantly valued F ratios. For the 3 task conditions the maximum correlation between efficiencies between-task was 0.17 indicating a remarkable lack of reliability. On the other hand the mean efficiencies (1.66, 4.35, and 2.17) for each control problem task, analysed for a single problem only, correspond to those reported in the literature.

The mean response-latencies were 25.3, 16.4, and 13.2 seconds for task 1 to 3. There is an obvious increase in speed in the course of the experiment. It seems plausible to assume that subjects spending more time in choosing their decisions should be more efficient in their performance. The data does not support this hypothesis. The correlations between average response latencies and the efficiency of performance are .03, -.09, and -.01.

Ray (1964) and Rapoport (1967) reported that approximately half of their subjects under-controlled the problems and half of their subjects over-controlled the problems. Similar proportions were observed for task 1 and task 3. But in the second task 69 out of 72 subjects exhibited over-control. Thus, also this proportion seems to be highly task specific.

(b) The Multi Stage Betting Game

The multi stage betting game resembles guessing about a sequence of throws of a biassed coin. The subject is required to maximise his score by winning points for successful bets. Once again, the experiment is computer controlled and two versions of the task were employed under 5 conditions of bias.

One version (free choice) presents the subject, at trial i, with a display showing his probability of winning, the number of stages remaining at the ith trial, and his score of points (for winning at previous stages in the experiment) at the ith trial. If the score at the ith trial is X_i, and Y_i is the amount bet at the ith trial (which may be chosen freely) then

$$X_{i+1} \begin{cases} = X_i + Y_i \text{ if win} \\ = X_i - Y_i \text{ if loss} \end{cases}$$

Five tasks were run with 15 stages each; the probability of winning was .6, .7, .5, .8, and .4. The tasks were administered in the same order to all the subjects and identical "random" sequences determined wins and losses. The other version (of the two major variants) relies upon the fact that under the assumption of a logarithmic utility function an optimal bet at trial i, if p is the probability of winning and if q=(1-p) is the probability of losing will be

$$Y^*_i = (p-q)\, X_i \quad .$$

The subject is here allowed to make a choice of bet by selecting from the possibilities

$$Y^* + 20\ \%,\ Y^* + 10\ \%,\ Y^*,\ Y^* - 10\ \%,\ Y^* - 20\ \%$$

Three tasks were run with ten stages each; the probabilities of winning were .6, .8, and .7. The sequence of wins and losses was the same as for the first ten trials of the free choice condition. In both conditions real money is paid to the subjects. The average amount paid to the subjects for both tasks, the control problem and the betting game, was approximately 10 $.

Some overall data are shown in Tab. 1, Tab. 2 and the average proportions bet under free choice and multiple choice conditions are shown in Fig. 1.

According to the optimal model, there should be a linear increasing relationship between the proportions which were bet and the probability of winning. From the averages given in Tab. 1 and Fig. 1 it is obvious that the actual bets are not in line with this prediction. Moreover, the optimal model predicts zero-bets for the unfair .4-condition and also for the fair .5-condition.

The behaviour is quite regular under the multiple choice condition, but in this condition the responses of the subjects were restricted to certain ranges. I also did some comparisons across the different tasks. An example is given in the correlations of Tab 2. These are inter-correlations with the bets made in stage 1. In stage 1 all subjects had the same current fortune, and it is possible to compare the data just for that round. For the other stages each subject has rather a different history of bets already made.

The intercorrelation for the free choice tasks are seen to be significant; though the correlation coefficients, are not very high. What is striking is the complete lack of correlation between the free choice and the multiple choice versions, although the task specificity is as high between slight modifications of task, as between the multiple and the free choice versions. Note the coefficients which are underlined in Tab 2. These have the same probability of winning.

The optimum policy predicts bets which are invariant in respect to three things. The percentage bet should be invariant with respect to the previous history, the current fortune and the number of remaining stages. The facts are shown in Fig 1, where plus and minus signs indicate wins and losses. Compare the four graphs and note, for example, the 0.8 condition - after losing subjects increase their bets whereas after winning they usually decrease their bet. This tendency in fact holds for the average proportions bet of 59 out of the 72 subjects. The contingency upon the preceeding outcome was not observed in the multiple choice condition.

In a pre-test a personality-questionnaire was administered to all subjects. The questionnaire (FPI, Fahrenber and Selg, 1970) in many respects is similar to the 16 PF and is one of the best known German test procedures. No significant correlations could be found between the nine scales of the FPI and any measures obtained from the control problem or the betting game.

We found a significant difference not in respect of measures related to the optimum policy, but in respect of response style. For example, there were significant differences between male and female subjects, explaining some 16% of total variance - female subjects, under multiple choice conditions, more often chose

probability of winning →		.4	.5	.6	.7	.8
mean bets for all stages	free choice →	47	53	60	41	49
	multiple choice →			24	41	59
mean bets for stage one	free choice →	35	48	57	59	60
	multiple choice →			27	42	61

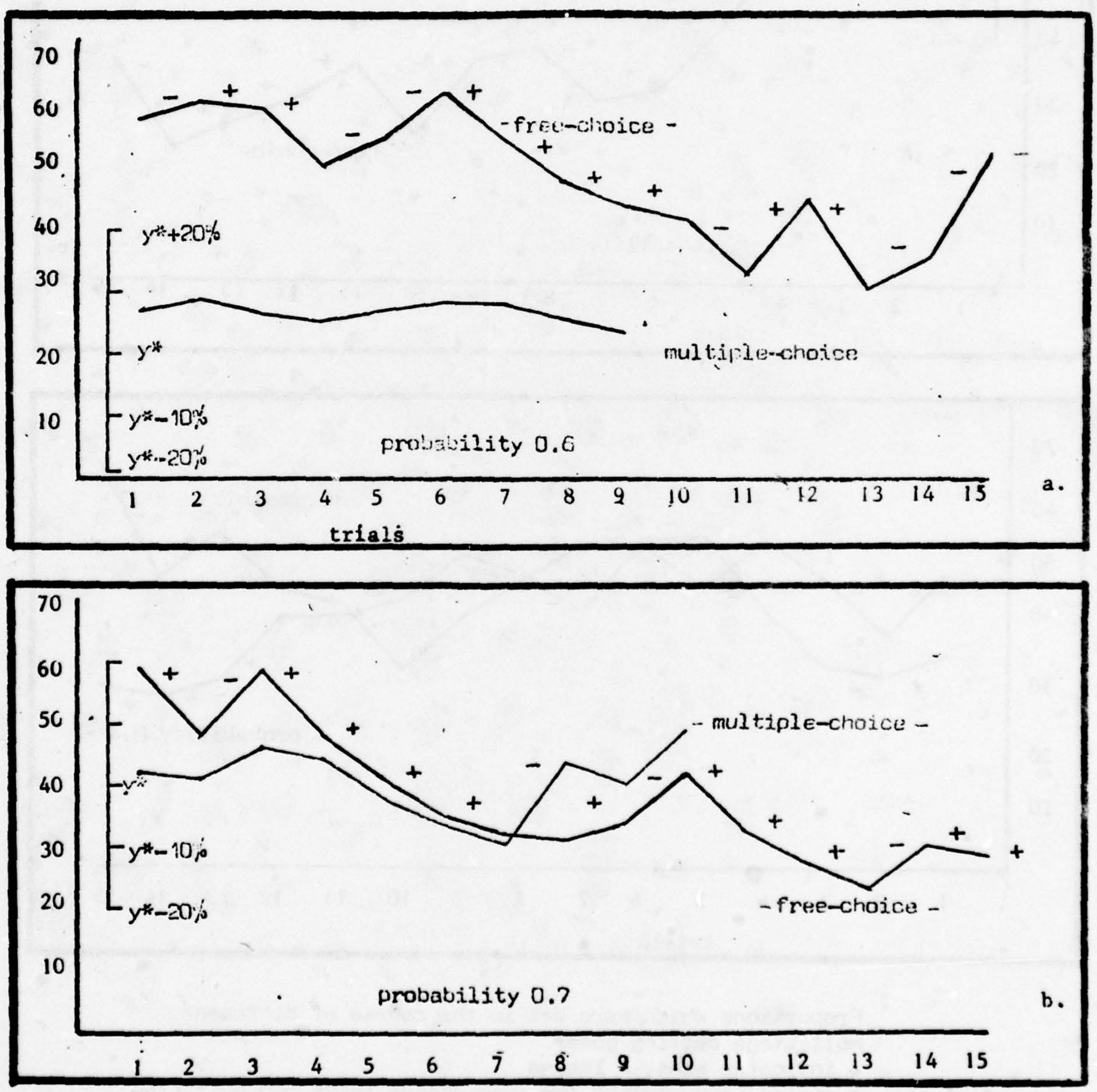
Mean bets for all stages and for stage one only in the different tasks of the *MBG*

Section 2.7. Table 1

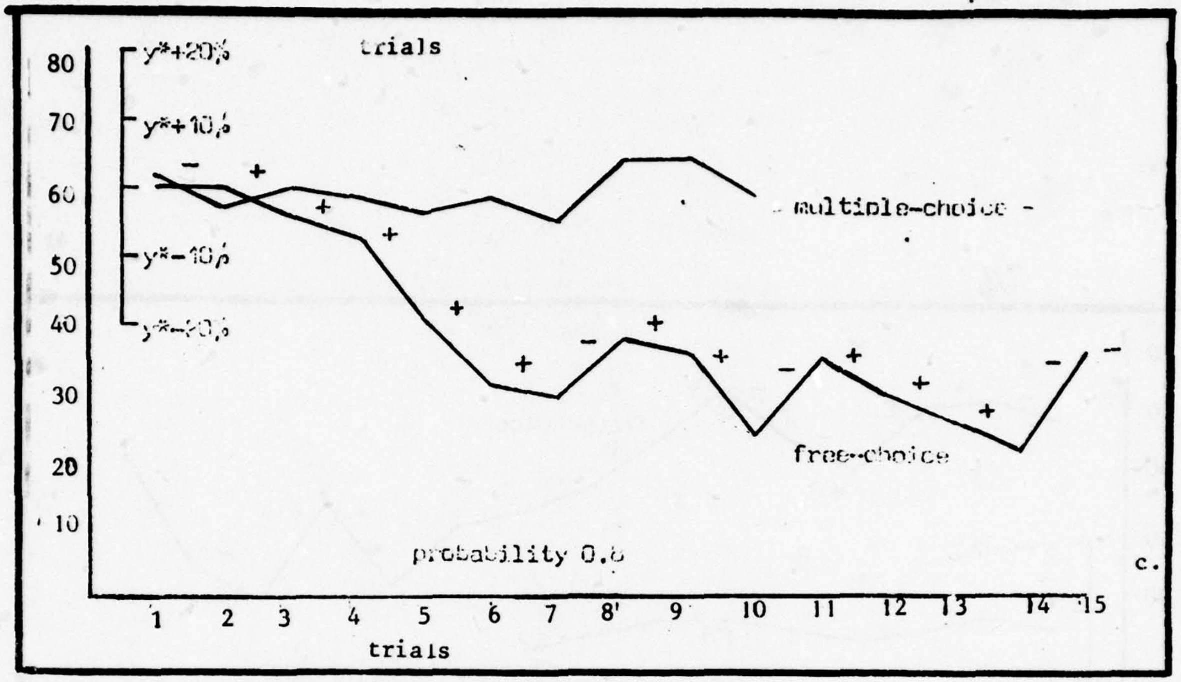
	free choice					multiple choice		
probability of winning →	.6	.7	.5	.8	.4	.6	.8	.7
task no.	1	2	3	4	5	6	7	8
1		39	43	37	37	16	14	20
2			60	53	47	15	18	-02
3				55	45	10	05	-04
4					53	24	14	17
5						21	10	22
6							40	33
7								22

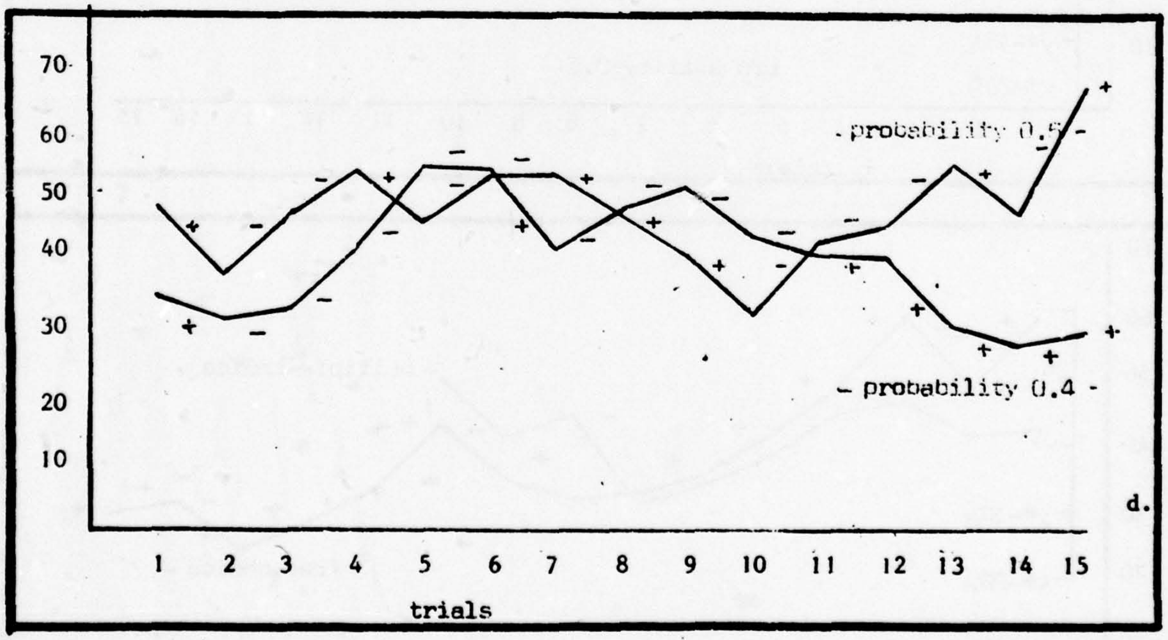
Section 2.7. Table 2

Intercorrelations between the bets of stage one; correlations for task having the same probability of winning under the free-choice condition and under the multiple-choice condition are underlined.



<u>**Section 2.7: Fig 1 (continued)**</u>





Proportions which were bet in the course of different multistage betting games
+ indicates wins, - losses
averages were plotted for free-choice tasks and multiple-choice tasks
y*+20%, y*+10%, y*, y*-10%, y*-20% were the alternatives under the multiple-choice conditions

__Section 2.7: Fig 1 continuation__

intermediate bets near to the optimal ones, and male subjects more often chose extreme bets, or under free choice conditions male subjects more often used round bets, ie. multiples of 5 or 10, and so on".

Discussion

Thus stylistic differences exist and, although the differences mentioned might be regarded as trivial, other differences revealed on careful scrutiny of response patterns (which are more difficult to quantify) are not open to this criticism. Kleiter next turned to some fundamental problems over the idea of optimality and an optimum policy.

For example, with the multi stage betting game a policy is only optimal where a logarithmic utility function is assumed, and not otherwise. The probabilities are just the probabilities announced to the subjects, not the subjects' personal probabilities as supposed. Many objective functions may be assumed but they are necessarily based upon assumptions made by the experimenter, not the subject.

"I am convinced that merely comparing actual behaviour with some optimum standard is too restrictive an approach, and a better knowledge of the cognitive processes involved in decision making is crucial for an improved understanding of multi-stage decision making. In contrast, it is promising to look out for general principles which are a basis for a more or less efficient control of the dynamic systems of our environment over the time dimension. For example 'how do we make use of recursive principles, upon which, dynamic programming relies so heavily?', or 'how do we process uncertainties of states in a network of inter-related states, as in a multi-stage decision problem?' or 'what is the influence of background knowledge about the imputation of constraints'.

For instance, in Broadbent's earlier example, of number guessing though no explicit constraints were mentioned, there were, in fact constraints about what was reasonable, what range of answers could be chosen and so on.

The transformation rule also mentioned by Broadbent, is, in my opinion, extremely important. Usually, the subject would not be presented with such a transformation rule, but would have to infer it from experience. This process of inference and finding the law of interaction between what the subject is doing over time (even if he does nothing) and what the environment is doing - is an important and also, as a rule, neglected, component in the psychology of judgement and prediction. At any rate, I hope that I have been able to present some material which casts some doubt on the rather restrictive and narrow approach of comparing actual behaviour to ideal standards."

Baker opened the discussion by broaching the question of response latency. There is quite a considerable body of work on latency (reaction time) as a function of the payoff obtained by the decision maker. For example, if the possible payoff is large then latency (or, depending upon whether the task is self paced or not) the time between trials is shortened, contrary to the current hypothesis. Conversely, latency is lengthened for certain cumulative

payoff assignments and the result is thus inverted and brought into line with Kleiter's proposals. As a rule latency depends upon the conditions and the decision maker's state of knowledge. Kleiter agreed but noted that most of these results depend upon the decision maker knowing (either explicitly, or implicitly by pacing) that latency is important and in his experiments the subjects were not informed that latency was being measured.

Pask tried to firm up the hypothesis at issue. "Your (Kleiter's) prediction is that if people deliberate over making decisions then the decisions are better made and the latency is longer". He said this precis was reasonable. "Most recent research has proved it the other way" responded Kahneman. "Perhaps, but the prediction is interesting", Kleiter said, "But the hypothesis was falsified" (so the results tally with most of the recent work). Baker and Pask insisted that it is still an interesting hypothesis. "Consider, for example, Kagan's work on the distinction between a reflective style and an impulsive style of response. The reflective thinker (or decision maker) ponders a problem deeply enough to pick out an unusual and often less likely alternative and does generally have a longer response latency as a result. Is that what you (Kleiter) had in mind; a distinct type of cognitive process?"

Kleiter replied "Yes, generating and checking up alternatives. Internal modelling and also, with these tasks, some kinds of calculation".

Turning to the issue of tranquilisation (one variable used to discriminate groups of subjects) Zeidner pressed for a statement of the original prediction. This was complex, but one prediction was that tranquilised subjects would have reduced risk prominence. In reply to Wendt, another prediction was an interaction between tranquilisation and the terminal score. But, in fact, none of these differences (nor any of the more complex types of difference) were found to exist in the results.

Kahneman spoke at some length on the "optimality" oriented approach". Here is a brief extract :

"What I fail to see is why these results are critical of the approach of defining an optimal strategy and comparing behaviour to it. What has been shown is that certain tasks, which are formally similar to one another, are not even models (comment under interpretation of the formalism) of one another. If so, then clearly one cannot expect behaviour in these tasks to generalise to other, more significant, behaviour. I fail to see what role the optimal policy has in your conclusion. I would take your results as rather a severe indictment on this particular type of experimental situation which is repetitive, highly artificial, highly contrived, fairly difficult for the subjects. The behaviour that you are eliciting is interesting and important. Notably with respect to the response to gains and losses, but the behaviours you are eliciting are not related to one another. Presumably, they are related to nothing else so the result is an indictment of the tasks, rather than of the optimality".

Zeidner said "Do you accept that conclusion?"

Kleiter said "Yes", without sounding too convinced about it, and proceeded as follows :

"For studies of other types of tasks, where we can control matters more thoroughly, we obtain similar negative results. One alternative would be to take more complex tasks, more realistic tasks, in which case the task becomes so complex that the optimality-oriented approach is restricted because optimal policies cannot usually be specified".

Pask agreed with the gist of the idea and asked Kleiter a question: "As I understand it, you are saying two things. First, the tasks employed in these studies are not complex enough to exteriorise the strategic and conceptual oddities that really distinguish styles of decision making, so that they can be observed as stretches of behaviour. On the other hand (as the next point of the two) if the task complexity and thus the potentiality for exteriorising man/machine or man/man interaction were increased, the task would be formally intractable. Is that right or not?"

"Yes", said Kleiter, with greater enthusiasm.

Kahneman pursued a related issue.

"We are touching on the problem of validity of method. When we are running experiments or constructing models of some situation in the real world, we do so because people take decisions under uncertainty for high stakes, and usually unique, non-recurrent, decisions without opportunity to recoup. We try to model this real world, in order to get a lot of data, by having people play recurrence games with very low stakes, in a highly contrived, artificial, environment, and perhaps the conclusion is that the decision process in that type of situation is unrealistic. Since risk taking does not have its real life connotation". Kleiter agreed. "But it is better to interpret some betting games, like multi-stage betting games, than to use a sequence of bets and not to analyse them like a multi-stage betting game, which very often has been done in the literature, i.e. to analyse the sequence of bets, as though each bet were virtually independent of the others".

Tversky interjected, "The real point is the motivation for task selection. The selection of these tasks was motivated primarily by the fact that there is an interesting optimal solution, and, given that, to see whether people could figure it out under limited information in a sequential game. I guess that the answer is that they can figure it out, which would suggest that we need to be more explicit and thorough in our consideration of the task selection".

Breaking away from this theme, several people (Baker, Kaplan and myself amongst others) asked whether introspective or retrospective reports had been obtained from the subjects: Kleiter replied that work of this kind is currently in progress and similar data is being gathered while working with multi-stage betting. The point was pressed since Kleiter had mentioned five especially important cognitive factors as worthy of attention and significantly influencing choice. It would be interesting to know how the subjects conceive, for example, what kind of machine is involved in the control problem whether subjects learn to manipulate the contraption as they might learn a tracking skill or whether they

entertain beliefs (and possibly "superstitions" or "spurious") beliefs, about the machanism. Kleiter said that subjects gave coherent but different reports of the visual or purely mechanical representations they have in mind.

Hogarth focussed on the deterministic component of belief, more or less in Broadbent's earlier (Section 2.4) sense. For example, "Do people in this (Kleiter's) situation ascribe regularities to some kind of deterministic rule generating the game. For example, when stockbrokers talk about the stock market they think they 'feel the market', that 'they have the system' and so forth. The same comment applies to betting and gambling".

"Yes", Kleiter replied, "especially for inventory control experiments".

"They believe that there is some kind of deterministic law between stock level at the last stage and ...."

"Yes and seasonal effects, and so on".

Hogarth asked Broadbent for his opinion. "Do your people in the business games (Section 2.4) imagine rules?"

"Yes, indeed they do. Sometimes they are right, too. For instance, in discussing whether to have a high advertising expenditure and large turnover, they do not do it so much on the basis of whether their hypothesis holds true in the world, but whether the man who wrote the program is likely to think it true".

Which brings the discussion back, rather neatly, to Baker's earlier comments (Section 2.2), on the style and imagination of SIMTOS program writers and software designers in addition to the style and imagination of subjects.

References

Fahrenberg, J. and Selg, H., Das Freiburger Persoplichkeitsinventar, FPI-Hand-anweisung, Gottingen, 1970, Hogrefe

Kleiter, G.D., Mehrstufige Entscheidungsmodelle in der Psychologie, Psychologische Beitrage, 1974, 1, 93-127.

Kleiter, G.D., Estimating the planning horizon in a multistage decision task, Psychological Research, 1975, 37-64.

Kleiter, G.D. and Wimmer, H., Information seeking in a multistage betting game, Archiv fur Psychologie, 1974, 213-230.

Rapoport, A., A study of human control in a stochastic multistage decision task, Behavioural Sciences, 1966a, 1, 18-32.

Rapoport, A., A study of a multistage decision making task with an unknown duration, Human Factors, 1966b, 54-61.

Rapoport, A., Dynamic programming models for multistage decision making tasks, Journal of Mathematical Psychology, 1967, 4, 48-71.

Rapoport, A., Research paradigms for studying dynamic decision behaviour, in: Wendt, D. and Vlek, C.A.J. (eds.), Utility, Probability, and Human Decision Making, Dodrecht, Reidel, 1975.

Rapoport, A. and Jones, L.V., Gambling behaviour in two-outcome multistage betting games, Journal of Mathematical Psychology, 1970, 163-187.

Ray, H.W. The application of dynamic programming to the study of multistage decision processes in the individual, Dissertation Abstracts, 1964, 4820-4821.

Shuford, E.H., Jr., Some Bayesian learning processes, in: Shelly II, M.W. and Bryan, G.L. (eds): Human Judgements and Optimality, New York, Wiley & Sons, 1974.

Toda, M., Algebraic models in dynamic decision theory, in: Vlek, C.A.J. (ed): Algebraic models in psychology, Proceedings of the NUFFIC International Summer Session in Science at "Het Oude Hof", The Hague, 1968, 105-160.

2.8. Dr D. Kahneman: Mathematical Theory of Decision Behaviour

Comment: Dr Kahneman and Dr Tversky gave a joint presentation entitled "Principles of Decision Making Illustrated by instances in the world at large". It happens that they divided the task and that Dr Kahneman overviewed a mathematical theory, with which both of them agree, whereas Dr Tversky commented upon real life examples of bias and misconception on the part of Decision Makers. His remarks are summarised in the (following) Section 2.9 and bear the brunt of psychological justification; they may be regarded as illuminating the cognitive processes that underpin the behavioural relations exhibited in the present paper. In the connection it is well worth consulting Tversky, A and Kahneman D. "Judgement under Uncertainty, Heuristics and Biasses". Science 1974 Vol 125 pp 1124-1132 and (same authors) "Belief in the law of small numbers" Psychological Bulletin 1971 Vol 76 No 2 pp 105-110.

"The theory of utility dominates the theory or risky choice. This is not a new story. It has been true for about three hundred years. The theory is currently used in a great many ways, both descriptively and normatively. It is used normatively, for example, in the context of decision analysis, where it prescribes a course of decision that people should follow if they wish to be consistent and to obey certain axioms that just about everybody considers reasonable and sensible. Utility theory is also the major descriptive theory of how people behave and make choices under uncertainty, i.e. that people behave according to expectation principles and the principle of expected utility.

When people violate the principle of expected utility, and they often do so, then explanations are sought, either in terms of the formulation of the problem, or in the demand characteristics of the situation, but the validity, the descriptive validity, of the model, seems to be quite generally assumed. Much of economic theory starts from the assumption that people behave under uncertainty according to the principles of utility theory, and, to the extent that psychologists have worked in this area, they have generally assumed that people behave according to expected utility theory at least to the first approximation, and they have relied on this assumption in a lot of the work that has been done, for example, in the measurement of utilities or subjective probability.

The major observation that led to the development of utility theory in the first place was that people are risk averse. For example, they reject fair bets and typically refuse to participate in a bet where, say, they have a 50 percent chance of winning 10 of something and a 50 percent chance of losing 100. This bet should be accepted if people behaved according to expected value. Utility theory was required to explain why people do so. It asserts that people follow the expectation principle, but that they apply it, not to the cash values defined in the problem, but to some psychological values; that is, to psychophysical variables, which in fact, correspond to money.

(a) If, as a first assumption, we suppose that there is an attitude to money, there is a psychophysical function for money, and that this psychophysical function is concave, (Fig 1 later) then risk aversion follows as a matter of course. It follows that if I have an equal chance of moving up, or moving down by the same amount on the wealth axis, that is, getting an increment of wealth or losing an equal amount of wealth, then the psychological loss which is associated with the negative change is greater than the psychological gain associated with the positive change. Hence, this proposition is unattractive if I follow the expectation principle, and I would reject the fair bet.

Notice that in this approach the whole concept of attitude to risk has been taken away. In utility theory the attitude to risk is replaced by an attitude to consequences - the concave function over the domain of monetary consequences is sufficient to derive the risk aversion behaviour that was supposed to be virtually universal for people, with one exception that has been very troubling to utility theorists (there are other exceptions, counter examples, and paradoxes but the one exception is outstanding). Why do people purchase lotteries? They spend millions of pounds purchasing lotteries. There is little that can be done, within the context of utility theory to explain that sort of behaviour.

(b) The second basic assumption of utility theory is that the utilities which matter in making a choice under uncertainty are utilities for final consequences. Prescriptively, people should look at these utilities. When utility theory is applied descriptively, it asserts that people do, in fact, consider options in terms of final consequences, i.e. a decision maker will consider his whole state of welfare, "what will life be like for him when the gamble is played". All the assets, "everything that a person has", is supposed to enter in the final determination of the utility of consequences, and, therefore, of the utility of the gamble.

Those are the two basic assumptions (a) and (b). They look rather innocuous, and we shall try to show that they are both quite wrong and that neither of them applies at a descriptive level.

In particular two major effects seem to run counter to both of these principles, the "certainty" and the "isolation" effects. The certainty effect describes the wide class of violations of the expectation principle. The isolation effect leads to violations of the second principle; that it is attitudes to final consequences which matter, or which dominate the choice between risky options".

These effects have been observed in studies of several populations using sample sizes of 100 or more respondents and appear to be reliably evidenced and reproducible.

Dr Kahneman sketched a theory believed to explain these effects and predict a few of their consequences. He and Dr Tversky tentatively

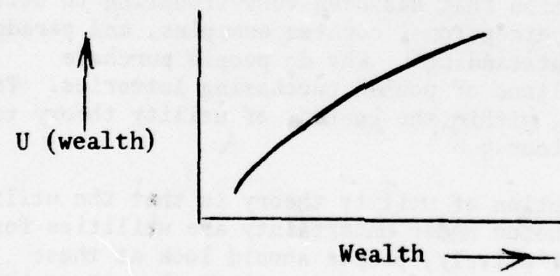
suggest this theory, as an alternative to utility theory, as a descriptive theory of how people actually make choices. The theory, as defined, is for two outcome gambles, but it is not tied to monetary gambles and it can be generalised, in a natural way, to more complex situations.

(1) The findings in support of the theory are based upon the questionnaire method. After careful explanations have been given and after any necessary discussion, the respondents are assumed to answer questions, mostly consisting in gambles or lotteries, with goodwill and without serious deliberate falsification. As a matter of notation, lottery options are represented in the form :

$$A = \langle x, p, y \rangle \qquad \text{or } B = \langle x^*, p^*, y^* \rangle$$

where x (x^*) is one option, p (p^*) is its probability, y (y^*) is the alternative option (since it is an alternative it has probability of $1 - p$, or $1 - p^*$). As a further convention, let U denote a utility of wealth gained, (where wealth is intended as a respondent's entire gamut of assets at the instant the gamble or lottery is presented).

(2) From standard utility theory, as applied to lotteries and the like, if the utility is a function, of the form shown in Fig 1.



Section 2.8. Fig 1

we may infer risk aversion (given $U(\langle x, p, y \rangle) = p(u(x)) + (1 - p)(u(y))$ and consequently that a gamble $\langle x, p, y \rangle$ is acceptable if and only if it is true that

$$U(\langle \text{wealth} + x, p, \text{wealth} + y \rangle) > U(\text{wealth})$$

This general statement is denied, in the sense that if fails to describe or predict actual behaviour (for example, the gambles that are accepted or preferred).

(3) First, there is a pervasive certainty effect which makes something sure loom very large. * For example, if the outcomes

* At the instigation of Stael von Holstein, Phillips and Hogarth it was explained that "certainty" is taken as "probability of 1". So far as small deviations (0.98 or 0.9) are concerned, the theory is designed to accommodate any reasonable departures from probability 1, and, at some stage, it would be necessary to accept that "near certain" departs from "certain" for example, probability 0.8" is perceived neither as "certain" nor even, at a guess, "near certain".

offered are A_1 = 100% chance of small holiday and A_2 = 50% chance of a very large holiday respondents prefer A_1 to A_2 (for notation $A_1 > A_2$). On the other hand if B_1 = 10% chance of a small holiday and B_2 = 5% chance of a very large holiday then $B_2 > B_1$. But, written out in the form $\langle x, p, y \rangle$

$$B_1 = \langle A_1, \tfrac{1}{10}, \text{nothing} \rangle$$

$$B_2 = \langle A_2, \tfrac{1}{10}, \text{nothing} \rangle$$

so that an inversion takes place, if both outcomes at issue are uncertain.

(4) If the outcomes are negative (regarded as detracting from overall wealth) then risk seeking is evident.

$\langle -1000, \tfrac{1}{2}, 0 \rangle \succ \langle -450, 1, 0 \rangle$ (ie. sure loss, of 450).

On the other hand, if the certainty effect is eliminated

$$\langle -450, \tfrac{1}{10}, 0 \rangle \succ \langle -1000, \tfrac{1}{20}, 0 \rangle$$

(5) By supposing a bonus, or given addition to wealth, it is possible to observe an isolation effect: namely, the bonus and the particular gamble are, at some point, disassociated. For example, if the respondents are told that they have a bonus of +1000 units then

$$\langle 1000, \tfrac{1}{2}, 0 \rangle \succ \langle 500, 1, 0 \rangle$$

But if the respondents are told they have a bonus of +2000 then

$$\langle -500, 1, 0 \rangle \succ \langle -1000, \tfrac{1}{2}, 0 \rangle$$

The two combinations of bonus and bet are identical when stated in terms of final assets. Thus, the preferences are inconsistent.

(6) Kahneman and Tversky seek to account for empirical findings such as these, and others, by a theory which is descriptive (ie. predictive of behaviour). Stated crudely, they wish to describe and predict behaviours that tally with the following empirical facts.

(a) For moderate values of p and positive outcomes, people avoid risk.

(b) For low values of p and positive outcomes, people are generally risk seeking (i.e., they gamble).

(c) For low values of p and negative outcomes, people are averse to risk taking (as a result of which, for example, they buy insurance as the converse of gambling).

(d) For moderate values of p and negative outcomes, people are risk seeking.

In fact the theory has rather wider predictive capabilities involving for example, mixed cases in which there is some probability of gain and some probability of loss.

(7) One key notion, is the idea of an uncertainty weight related to the asserted probability of an outcome.

"The uncertainty weight is not a subjective probability. I must make that very clear. What we are doing here, and this is a methodological choice, and a conceptual one, is to say that there might be better definitions for a subjective probability than looking at its weight in gambles. If, for example, I get somebody's evaluation of a gamble on tossing a coin, I think that for any reasonable person, if he believes that the coin is fair, the probability that it will land heads rather than tails is one half. But we want to conceptually distinguish the weight of that event in an uncertain choice from its probability. So we retain the concept of subjective probability, but do not define it by choice between gambles. In choosing between gambles something else is done and another weight is applied, an uncertainty weight which is a function of the subjective probability".

This uncertainty is related to the numerical value of the asserted probability (p) and may thus be represented by an uncertainty weight π (p) with the following general properties, exemplified in Fig 2 namely :

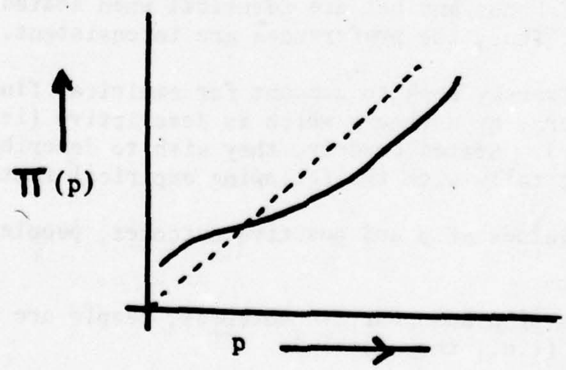
(a) $\pi(p)$ increases with p

(b) $\pi(p) > p$ for small valued p

(c) $p > \pi(p)$ for medium or high valued p

(d) $1 > \pi(p) + \pi(1-p)$ (which reflects the certainty effect)

(e) $\pi(o) = 0$ and $\pi(1) = 1$



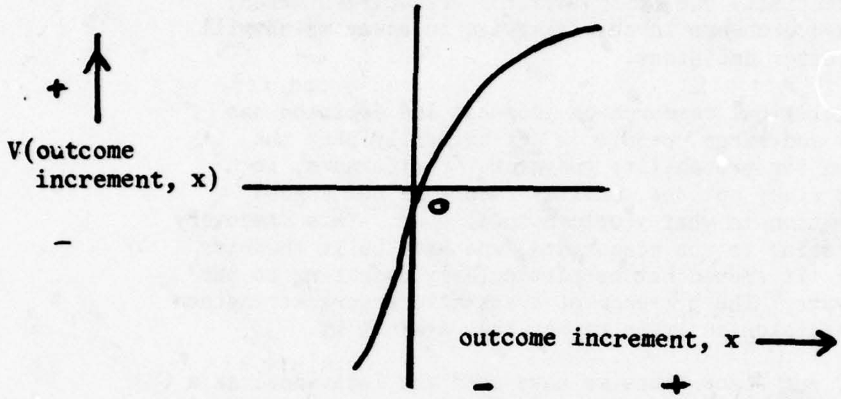
Section 2.8 Fig 2

The uncertainty weight π (p) is combined with a value function (V) specified only for changes in wealth (that is, for outcome increments positive or negative) which has the general properties exemplified in Fig 3.

(a) Concave above 0 (positive increment)

(c) Convex below 0 (negative increment)

(c) Convex slope steeper than the concave slope and v = 0 at the origin



Section 2.8 Fig 3

"Our view of the value function follows directly from the isolation effect. By and large people do not evaluate uncertain options in terms of the final consequence, that will be obtained when the gamble is played, they define and look at uncertain options in terms of change, in terms of gain or loss. So, if we want a descriptive psychological theory for the way that choices are made, we must describe these choices in terms of gains or losses, and not in terms of final consequences, only".

(8) Returning to the $\langle x, p, y \rangle$ notation, the theory may be summarised by a pair of statements predicting a respondent's choice in gambles and lotteries, as follows: assuming that such preference choices are based upon perceived (π weighted) probabilities and perceived (V) values

(A) Unless $x > y > 0$ or $0 > y > x$

$$V(\langle x, p, y \rangle) = \pi(p) \cdot V(x) + \pi(1-p) \cdot V(y)$$

(B) If $x > y > 0$ or $0 > y > x$ then

$$V(\langle x, p, y \rangle) = y + v(\langle x - y, p, 0 \rangle).$$

The chief protagonists in a brief discussion were Keeney and Stael von Holstein. Efforts were made to reconcile the theory with either "standard" or multiattribute utility theory. It appears that a complete reconciliation is untenable, at the moment, and is likely to be impossible because the current theory embodies the well founded observation that subjects evaluate the risky component of a situation and the other than risky components, differently. So far as particular cases are concerned, compromise is possible but Kahneman and Tversky have tried most of the general types of "standardisations" that were suggested and found them, for one reason or another, inadequate.

2.9. Dr A. Tversky "Interpretations, Theories heuristics and Propensities.

Apparently, there is an agreement that decision analysis which consists of the expectation principle, the probability calculus, and utility theory, provides an adequate language for the analysis of decision-making in an uncertain world. The problem is essentially one of application or implementation; how to structure problems in this language in a way which will help us make better decisions.

The psychological research on judgment and decision has shown that, by and large, people do not naturally obey the rules laid down for probability judgment. Furthermore, in choice between risky options, utility theory is not a good first approximation to what people naturally do. This discovery should be upsetting to the economists, who have built theories on that basis; it should not be particularly upsetting to the decision analysts. The presence of systematic errors strengthen the case for decision analysis rather than against it.

In all of our discussions we have used the individual as a measuring instrument. The decision analyst records measurements from a human being who is generally an expert regarding the problem under study. Thus, he is involved in the measurements of beliefs and of values. But decision analysis is a theory of measurement without serious consideration of the measurement device itself, i.e. the human judge.

Dr Tversky went on to develop an alternative approach to decision analysis which differs subtly and significantly from either the normative or the descriptive paradigms as they are usually envisaged.

Though an agreement reflecting a compromise may often be achievable (perhaps by dint of the decision analyst acting as an educator, on which point Dr Tversky has several reservations) there are occasions upon which decision makers are apparently unable to change their way of thinking to suit the mores of any rational decision theory. This seems to be true even of professionals trained as probability and utility theorists (like many of the subjects in Broadbent's and Kahneman's experiments). The characteristic habits of thought are deeply ingrained or even innate* and Dr Tversky concentrated upon one of them in expanding a "propensity interpretation" or "causal interpretation" of the probability of joint and conditional events.

---

* Comment: For example, Tversky and Kahneman have previously noted a fairly ubiquitous heuristic, representativeness which involves a confusion between similarity and probability. The heuristic may influence very many types of judgment: for example, the estimated profit of a firm is biassed by existing descriptions of the firm: if the firm is similar to a successful enterprise, the profit is likely to be over-estimated for that (not for a probabilistic or statistical) reason. Short sequences of random events are taken to be representative of a larger sequence, or a non existent event generating system may be assumed (superstitious assumptions about the generation of random stimuli in a guessing and betting experiment) again as a result of a similarity judgment. Finally, performances (of tasks and the like) are taken as representative of the performer, ie similar to his typical performance, so that the effects of regression (towards the mean) are frequently underestimated or neglected.

"There was a statistic like that (Zeidner) in Washing DC. Some 95 percent of purchasers of burglar systems buy them after the first burglary".

"Let us now examine the probability of burglary during 1975 given that a lock has been bought. In this case people agree that the probability of burglary, given that the lock has been bought, is less than the probability of burglary before the purchase of the lock.

"People believe that burglary increases awareness of the problem and leads to the purchase of locks; conversely, the purchase of locks prevent burglary".

"We do not all think about the problem in the way that you are posing it (Philips). I would never ask a person to answer such fancy questions if he suddenly asked about time sequence; the decision analyst should say 'He is thinking about it in time-sequence terms'. I had better decompose it".

The probability that the lock is bought, conditional on a burglary is greater than the probability of buying a lock. But there are two cases: either the break-in occurred before the lock was bought, in which case the likelihood of purchase is increased, or it was bought after the burglary, in which case it did not matter.

Keeney, Hogarth, and Von Holstein debated several interpretations of the problem before Tversky countered:

"Almost everybody that we have studied, and we tried this on some very sophisticated people, violated the precepts of probability theory. The probability of buying a lock given a burglary is denoted $P(L \mid B)$, and is equal to the probability of L <u>and</u> B, divided by the probability of B. Therefore, the statement $P(L \mid B) > P(L)$ is equivalent to the assertion that P(L <u>and</u> B) is greater than the product $P(L) \times P(B)$. The joint occurrences have greater probability than the probability of their independent occurrence. Placed in a two-by-two table, there is a positive correlation in the sense that L and B are more likely to occur together than one would predict from multiplying their marginal probabilities $(P(L) \times P(B))$.

The statement $P(B \mid L) < P(B)$, on the other hand, implies that P(L and B) is less than the product $P(L) \times P(B)$, which is, of course, impossible! These two judgments could not be maintained for one implies the opposite of the other.

The above inconsistency stems from the tendency to interpret conditional probabilities causally. It is close to impossible, however, to use our own knowledge to elicit sensible conditional probability estimates, except by thinking about a situation in an implicative or in a causal sense.

If we really want to determine conditional probability in the strict sense; and, of course, all probabilities are basically conditional (Comment: upon the structural hypothesis always, even from a frequentist point of view. More generally, upon the subject's prevailing condition of belief and state of information) it would be necessary to ask is the probability of burglary and buying a lock greater than or smaller than the probability of a burglary times the probability of buying a lock. The logical lesson is that in order to achieve this result problems need to be decomposed, or reformulated, to remove the time element. One should be extremely cautious in interpreting judgments of conditional probabilities which are generated through causal model, as proper conditional probabilities.

Von Holstein interjected , "In real life, where everything is conditional - any real value is conditional on so many different things.

Instead of trying to force a conditional probability or probability calculus on the subject, we could build a model of his thinking about the problem, which may be a causal model. In this model, estimated conditional probabilities are viewed 'propensities', or measures of the strength of causal relations between variables.

In this approach we search for the most appropriate mode in which the expert can communicate with the analyst. What is the mode that is most appropriate for him to convey the information that he has? It depends, probably, on his way of thinking.

This approach is part of a general effort to develop a technology for decision analysis, which takes into account not only the formal theory, but also the characterists and limitations of people and systems".

Kaplan asked if anything of the kind had been tried in practice or in the laboratory.

"I have made a case for an alternative way of thinking about the problem. I am not sure that it is the preferred alternative. We have made preliminary attempts to develop decision procedures along these lines".

"So, (Kaplan) there is a commonality of view between you and Broadbent".

"Probably there is. We are trying to bring psychology into decision analysis by considering the psychological limitations of decision makers".

Baker brought up a fresh point: "Something I have often wanted to do, but have never done is as follows. Suppose one asks a subject the probability that so-and-so will happen, and he says P = 0.6. I wanted to ask him afterwards; 'now, if you have to make a bet, what is the probability that your assertion that P = 0.6 is correct'. Suppose he says P = 0.75. Next, given a bet that pays off on that bet, see where you reach P = 1 as a means of delineating what that P = 0.6 really means".

"You have touched on an important problem; the confidence interpretation of probability (How <u>sure</u> are you that your probability is <u>correct</u>?). Confidence judgments follows somewhat different rules that violate the normative theory in yet another way. We should be aware of the multiplicity of interpretations if we want to make use of the information given to us".

Keeney argued in favour of a mixed approach. "The (descriptively oriented) model that you are suggesting and, say, a decision analysis model. That is how I try to do it. In the process of interviewing somebody, if I can get their utility function, I can get a fairly good idea of what their real preferences are. They tell me some things that violate what I consider to be reasonable consistency criteria. I certainly discuss rationality with them and denial. If I get to know them well enough, I feel that I can perhaps match their actual utility function, and some of their actual probabilities, too".

"We should go for a major, massive educational effort, to get people straight about the foundations of probability. It is quite difficult to discover inconsistency unless one knows exactly where to look for it. In this respect, the work on the certainty effect and the isolation effect that Kahneman described could be potentially useful for decision analysis in terms of knowing what to check for".

The remaining discussion concerned value and utility estimation. Many participants were involved but the dialogue, though lively, was not as coherent as the rest. Moreover, most of the issues had either cropped up previously or are stated more clearly in later sessions.

2.10. Dr R.M. Hogarth: "Toward an Idiot's Guide to Decision Making: or on the futility of 'optimality' "

"An important task for decision analysis is to determine the structural characteristics of decision situations where man can be taught to use simple, but analytically sound decision rules or heuristics. These need not provide "optimal" solutions, although they may be optimal if the formal cost of decision analysis is considered. The key idea is to examine, analytically, the properties of decision situations so that easy to apply rules of thumb can be derived for use by the practitioner. In practice, this subsequently requires that appropriate decision situations can be easily recognised; and that the rules of thumb can, indeed, be easily applied. There is a need for an idiot's guide to decision making, especially in the context of an apparent dissatisfaction with the practical usefulness of the management and decision sciences".

Dr Hogarth illustrated his thesis by noting that some commonly but often covertly employed expedients (like assuming uniform weighting of coefficients in a linear combination of variables) could be justified on formal rather than intuitive grounds and thus recommended as heuristics, provided users are alive to the limits of legitimacy. However, most of his argument was concerned with the question of group judgement, "when $n \geq 2$ heads are better than one head" and with the manner in which the judgemental "heads" should be aggregated (as by consensus, averaging, or choosing one judge or decision maker of superior ability and using his choice as decisive).

Here, also, a formal analysis is possible and leads, at least, to a baseline index with which to compare judgemental performance, possibly achieved by other methods (for example, group discussion, or coherence based agreement between the participants). Since the mathematical issues are presented thoroughly in a paper (Einhorn H.J., Hogarth R.M. and Klempner, E. "Quality of Group Judgement") only the conclusions are summarised and some supportive data presented in a condensed form. In manufacturing a precis I have deliberately emphasised the assumptions underlying the argument in "two Heads" because they are inherently interesting and were most avidly discussed. To facilitate cross reference the notation used in the source paper is preserved intact.

The following group decision strategies are considered. Their relative merits, for different groups sizes (N) and for different true values (X_t) of a judgemental variable, X, are compared, in terms of E(d), the expected deviation between the predicted value of X and its true value X_t, in Fig 1, Fig 2, Fig 3, Fig 4. These strategies are (1) Random, meaning that one person is selected at random from an N person group and, supposing this is the j^{th} person ($j = 1, \ldots N$), that his predicted value of X, namely, X_j, is used as the group judgement (2) Mean: in this case, the mean value $\bar{X}_N$ of N distinct judgements is used as the group judgement (3) Best: meaning that some method exists for identifying the one member of the group who can best estimate the true value of X and thus his individual judgement X_i is used as the group judgement, where the i^{th} individual is best ($i = 1, \ldots N$). (4) Proportional: There is a probability $p_{i(N)}$, of selecting the best member, rather than certainty in the matter, and it is assumed that this probability is inversely proportional to each

members' ranking of liability to deviate in judgement from X_t. This scheme is clearly identical to the random scheme ($p_{i(N)} = \frac{1}{N}$) for no such rank ordering and is chosen as a "conservative" example of a scheme with differential weighting having some validity.

The main pattern of the results is determined on scrutiny of the graphs in Fig 1, Fig 2, Fig 3, Fig 4 (for example, that random gives the highest expected deviation, and best the lowest, and the crossing over of E(d) proportional and E(d) mean that takes place between values of $|X_t| = \frac{1}{2}$ and $|X_t| = 1$.) Supposing that the utility of a decision (U) decreases with d as a linear function like:

$$u = a - b\ .d$$

and that we wish to maximise the expected utility

$$E(u) = a - b\ (E(d)\ ).$$

these graphs alone lead to some rather clearcut recommendations in favour of the proportional strategy (rather than the mean strategy). The random candidate is out of court, because it is, in all cases inferior, the best because it is probably unachievable in practice.

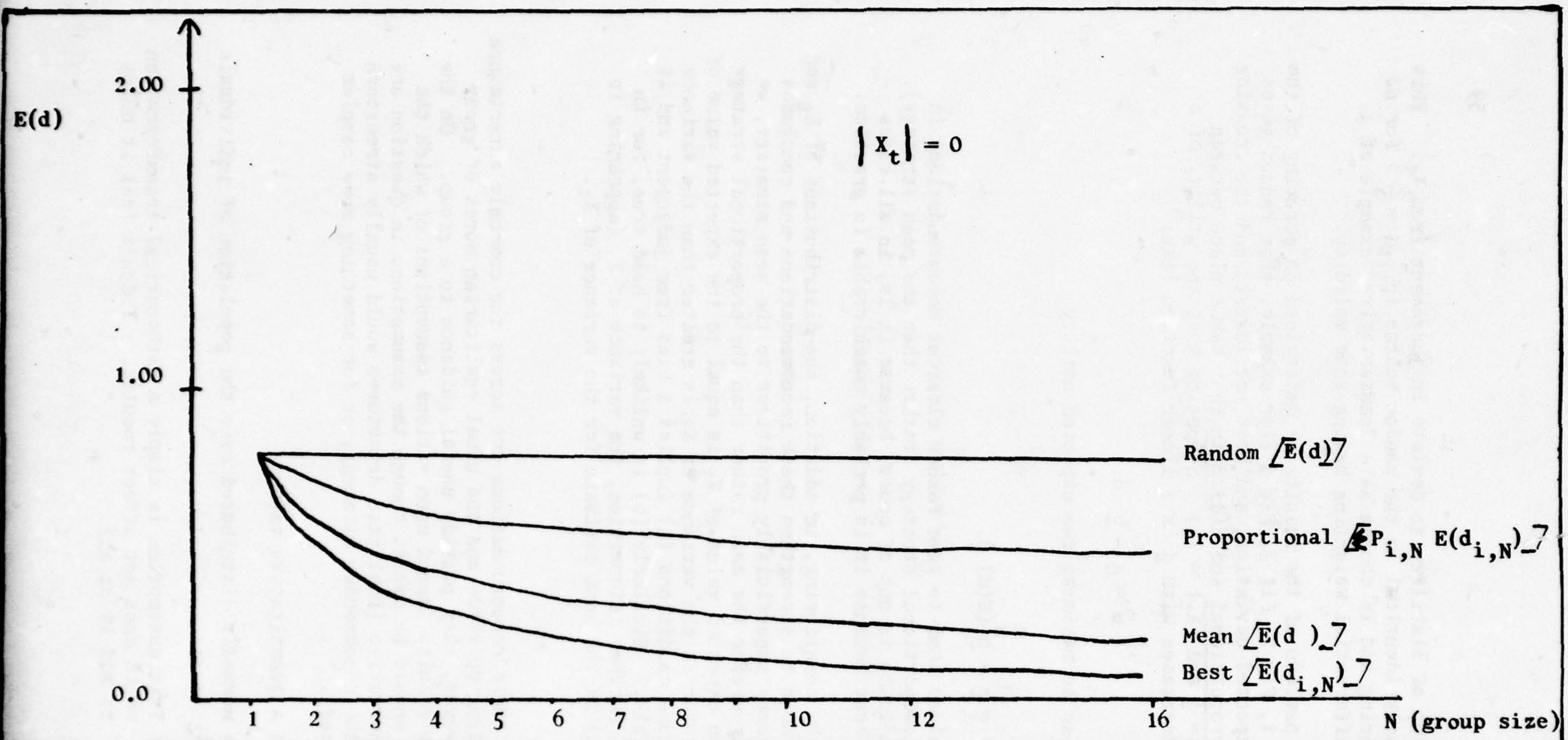
However, by considering, in addition, the distributions of X_t and X_j Hogarth is able to strengthen these recommendations and concludes that even in cases superficially propitious to the mean strategy, we should actually prefer the mean rather than the proportional strategy only if (a) the expected value of X_t is equal to the expected value of X_j and (b) insofar as the variance of X_j is greater than the variance of X_t. Of these conditions (a) implies a bias free judgement and is often implausible. Similarly (b) is unlikely to hold true, for in the absence of further information, the variance of X_j (appearing in condition (b)) is the best estimate for the variance of X_t.

Clearly, these recommendations cut across the commonly entertained notions of adding up votes and the usual egalitarian mores of group decision: however, they provide useful guidance to a group. On the other hand the results depend upon various assumptions of which the group or user should be aware, though the assumptions in question are generally conservative (realistic departures would usually strengthen the case for the recommended strategy, or for something more complex and personalised).

(1) X has a quantitative value

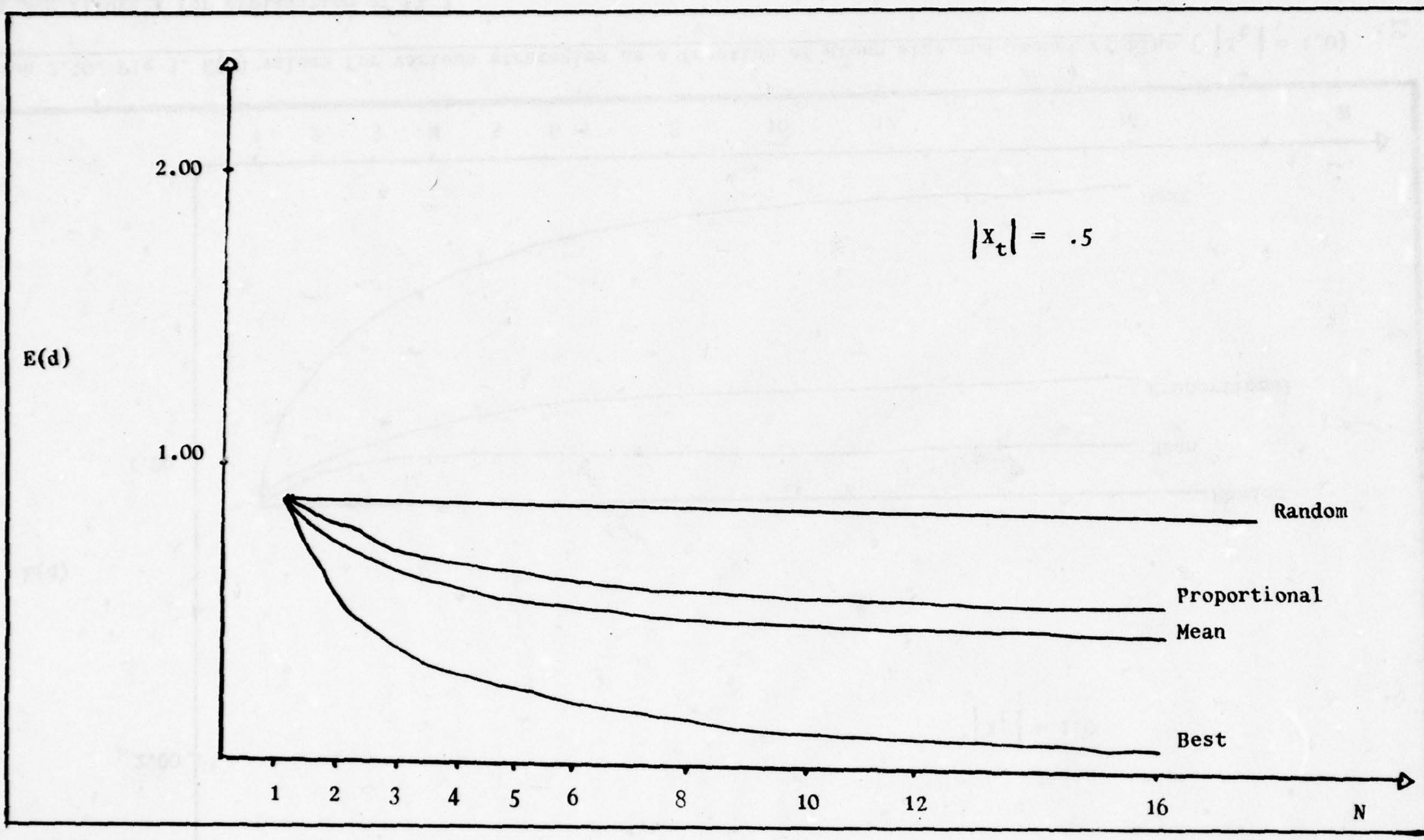
(2) X_j is normally distributed over the population of individuals $j = 1, 2, \ldots$

Note: This assumption is simply a mathematical transformation which does not affect results. I don't feel it helps to add it at all.



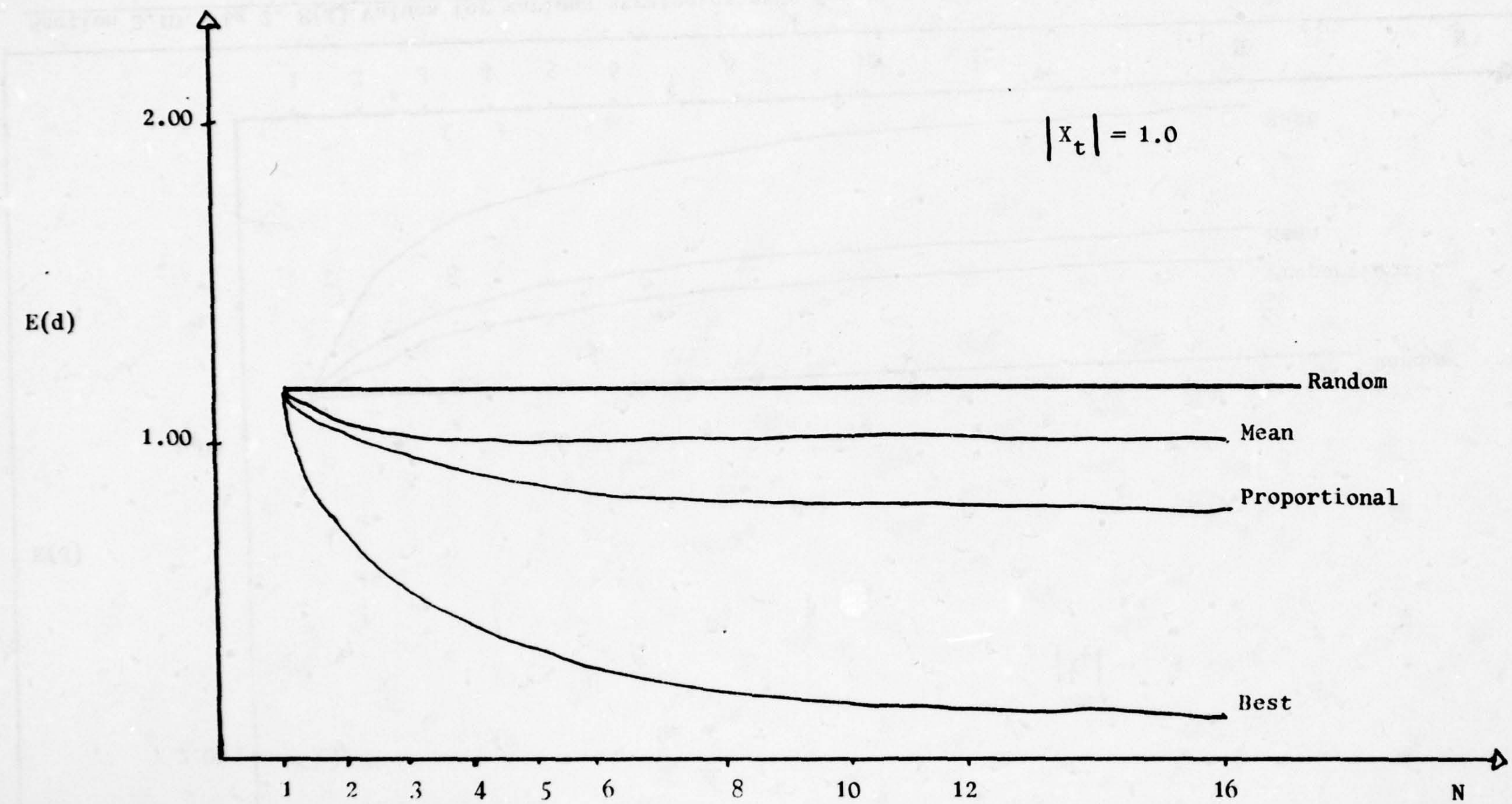
Section 2.10. Fig 1 E(d) values for various strategies as a function of group size and amount of bias ($|X_t| = 0$)

Note: $|X_t|$ is the absolute deviation of true value to be predicted from the mean judgement of the population of experts from whom N are chosen. It is measured in units of the standard deviation of the distribution of experts' judgements.



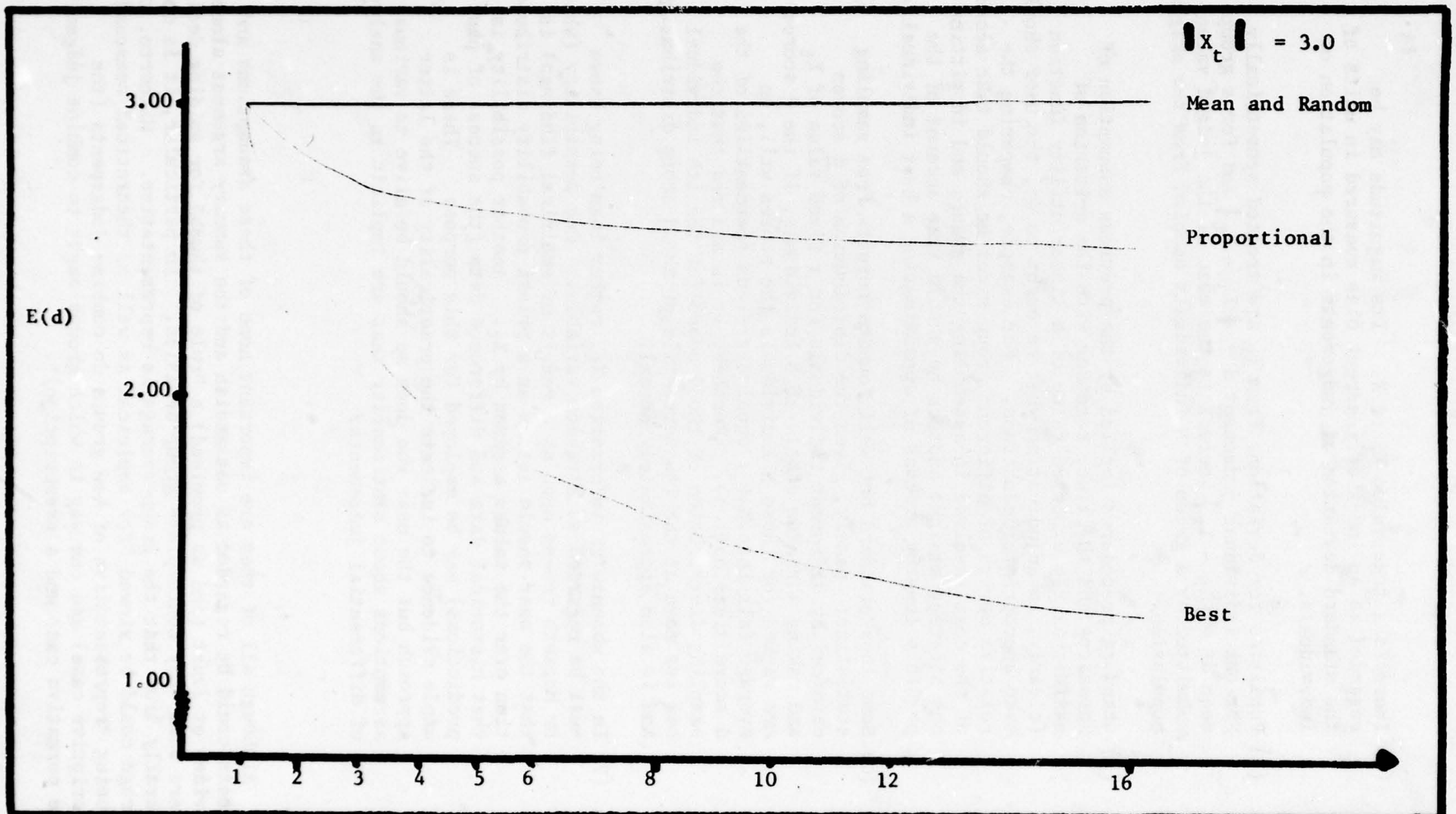
Section 2.10. Fig 2. E(d) Values for various strategies as a function of group size and amount of bias ($|X_t| = .5$)

Note : See figure 1 for explanation of $|X_t|$



<u>Section 2.10. Fig 3. E(d) values for various strategies as a function of group size and amount of bias</u> ($|X_t| = 1.0$)

<u>Note: See figure 1 for explanation of</u> $|X_t|$



Section 2.10. Fig 4: E(d) Values for various strategies as a function of group size and amount of bias ($\|\bar{X}_t\| = 3.0$)

(3) There is a true value X_t of X. Its magnitude may be regarded as an index of inherent bias measured in units of the standard deviation of judgements in the population of individuals.

(4) Penalties for deviations from X_t are treated symmetrically. For one individual judgement $d = |X_j - X_t|$ and for a group mean $d^* = |\bar{X}_N - X_t|$ where $\bar{X}$ is the mean of the judged values submitted by a group of N individuals sampled from the original population.

(5) Standard procedures implied by the previous assumption of normality and the like, together with the criterion of maximising the expected value of a linear utility function (clearly, in adjudicating what he ought to do, the user should make appropriate deductions. For example, comparing the relative merits of different group sizes he should take account of the cost involved in assembling the group, and in pitting one strategy against another he should take account of the possibly lengthy process of approximating a best individual).

(6) Some less standard but well founded results from sampling statistics; namely, given the distribution of d scores obtained by different individuals for a fixed value of X_t and taking a random sample of N individuals, if the d scores are ranked for these N individuals the scores will, on average fall into N + 1 equal divisions (percentiles of the d score distribution). Further, it is assumed that the sampling distribution of the d score for the ith individual has its mean at the ith percentile of the d score distribution and is also approximately normal.

(7) In the absence of information X_t, rather than being known must be regarded as a random variable. One possibility (which Dr Hogarth frowns upon as a result of empirical findings) is that the user should assess an a priori probability distribution over the values assumed by X_t. Another possibility is that historical data and difference data (the success of past predictions) may be employed for this purpose. There is ample evidence to indicate the practicality of the latter approach but the user who does so should be alive to various assumptions about stationarity that are implicit in the analysis of differential judgements.

Although all of them are important none of these assumptions and cateats could be regarded as outlandish and the summary argument alone provides at least (and as promised) a "rule of thumb" for guiding decision makers when they contemplate group decision. In particular, it is not generally true that the group average is representative. Hitherto, the average could be viewed "for empirical as well as theoretical reasons" as being "representative of how groups do combine judgements (the descriptive case) and one way in which groups ought to combine judgements (the normative case and a prescription)".

Dr Hogarth's presentation aroused quite lively discussion, only some of which took place around the table in the room. As an interesting comment upon the group process, nobody, on the previous day, was prepared to defend the Delphi forecasting technique (which relies upon systematic averaging) whereas, on this occasion, there were advocates of its more sophisticated variants (Zeidner, Wendt, Von Holstein).

It was stressed that Dr Hogarth's method provides a comparative measure for interpersonal interaction. "Considerable theoretical and practical effort has gone into the idea that a group decision is a synergistic product. To determine the truth of this hypothesis one needs, at least, a comparative baseline. Such a baseline might be provided by the Best Person strategy, synergy being demonstrated insofar as the performance of a group can surpass the Best Person Strategy level".

Another use of the scheme is in "guiding research into how (by controlled group discussion methods, for example) the group can select best members. The groups effectiveness in selecting best members can be measured and represented in terms of the selection probabilities p_i" (so far as these deviate from the p_i that are assigned by the proportional and the random schemes).

2.11. Dr L. Phillips: Individual and Cultural Differences in Assessing Probability

"My main focus of discussion today will be the question of how people think and react when they are faced with an uncertain situation; perhaps even before one asks them to encode probabilities. The investigations are partly motivated by experience in teaching decision analysis over a period of several years, and finding that some students feel quite violently opposed to, or have great difficulty in understanding the concepts of uncertainty, utility, etc. This experience led me to examine not only individual but also cultural differences in the way that people face up to uncertainty. Perhaps our Western way of thinking imposes a particular structure that has been embodied, subsequently in decision analysis. Apart from that, there is the general question of how cultures differ from one another in their regard of uncertainty".

Dr Phillips expounded a basic cultural hypothesis: that some cultures (notably, the Chinese) are fate-oriented whereas others are achievement-oriented (Japanese, at any rate to some extent, the Western cultures). A member of a fate-oriented culture acts in the belief that things are predetermined and is unlikely to maintain probabilistic hypotheses; vice versa, achievement-oriented people may plausibly do so. Hogarth suggested, some years ago, that subjects from fate-oriented cultures would score highly on "external control" (from outside forces) and those from achievement-oriented cultures highly on "internal control" (they ordain their own actions) using Roter's scale. He confirmed this finding at the meeting and Phillips own studies have led him to conclude that such a distinction exists though more dimensions are required to give a full account of the intercultural differences. Specifically, a literative survey strongly suggested :

"The Americans are more internal; the Chinese born in America are more internal than Hong Kong born Chinese; Indians are very internal and, at the other extreme Japanese are very external. Israel, France, Canada, Germany and the U.S.A. are all somewhere in the middle".

The other cultural dimensions, of relevance to uncertainty, are authoritarianism, tolerance of ambiguity, rigidity, conservatism and dogmatism. There is evidence that some cultures appear more authoritarian than others. But authoritarianism arising from social mores has to be distinguished from the personal variety.

"When trying to find the way a person views the world, we did not want to plunge right into probability assessment, since we felt that we might be imposing an unfamiliar measuring procedure upon the subjects. Instead, we developed a 'view of uncertainty questionnaire'. (Fig 1). The aim of which is to determine how an individual views the world in probabilistic terms. Respondents are asked to give a reasonable and appropriate response to the questions.

Responses are coded into the following categories: yes or no; don't know; a 'probability response' (perhaps, possibly) a conditional yes or no (yes if), a conditional probability; and a catch all category. The categories are determined by pilot studies and a

20. Will a cure for cancer be discovered before the year 2000?	20.
21. Was the Ancient Mayan Empire located in Peru?	21.
22. Was paper invented by the Egyptians?	22.
23. Will a television set so flat it can be hung on a wall ever be mass-produced?	23.
24. Will the Pope ever sanction the pill?	24.
25. Was Thomas Jefferson the 3rd president of the United States?	25.
26. Did Tanzania have any TV stations in 1973?	26.
27. Will man set foot on Mars before the year 2000?	27.
28. Will Britain ever formally grant independance to Rhodesia?	28.
29. Will South Africa be banned from the next Olympic Games?	29.
30. Is the Nile the world's longest river?	30.
31. Did Caesar invade Britain in 55 BC?	31.
32. Will a woman every become Secretary General of the U.N.?	32.
33. Has South Africa the highest waterfall in the world?	33.
34. Will Germany ever become reunited under a single government?	34.
35. Do both the Tigris and the Euphrates rivers empty into the Red Sea?	35.
36. Is Copenhagen more than 500 miles from London?	36.

space in which subject is to construct reasonable and appropriate responses to questions

Section 2.11. Fig 1.

1.	Which metal is denser?	(a) Iron (b) Copper	______	Space for subject to construct a % sure numerical response maximum at 100% minimum 50%
2.	Which contains more proteins per unit weight?	(a) Eggs (b) Steak (Beef)	______	
3.	Did the USSR and Denmark	(a) Join the UN at the same time, or did (b) The USSR join first?	______	
4.	Which has the greater average life expectancy?	(a) Elephant (b) Man	______	
5.	Which is larger?	(a) Black Sea (b) Caspian Sea	______	
6.	Which is the second official language of Finland?	(a) Norwegian (b) Swedish	______	
7.	Aleksandr Solzhenitsyn was awarded the Nobel Prize for Literature for which year?	(a) 1970 (b) 1971	______	
8.	How did Henry VIII get rid of Anne of Cleves?	(a) Had her beheaded (b) Divorced her	______	
9.	Buddhism originated in which country?	(a) China (b) India	______	
10.	Which is further North?	(a) London (b) New York	______	
11.	The principle language of Brazil is	(a) Portuguese (b) Spanish	______	
12.	How did Mussolini die?	(a) Suicide (b) Execution	______	
13.	The egg of which bird requires longer incubation	(a) Chicken (b) Pigeon	______	
14.	In 1963, Valery F. Bykovsky orbited the earth in Vostok V. Was this cosmonaut	(a) A man, or (b) A woman?	______	
15.	Which planet takes longer to go round the sun once?	(a) Earth (b) Mars	______	

Section 2.11 Fig 2

scoring scheme exists. For example, if someone says 'I hope so', this is not deemed an expression of uncertainty, or knowledge, but an expression which seems to have nothing to do with the analysis, i.e. an avoidance of the issue.

It should be noted that half the questions relate to future events and the other half to events where people could possibly know the answers. About half a dozen dependent variables are derived from the questionnaire responses and one of them is the number of correctly answered questions amongst those questions for which we knew the correct answer.

It is a simple questionnaire, but the interesting thing is that at least five to ten percent of people never use a probability response. They say yes, no, I don't know, and may use a conditional yes or no, but there is no probabilism in their response at all.

Some simple hypotheses are 'more yes/no's and fewer probability words from Chinese subjects than English subjects' and 'more yes/no's and fewer probabilities from nurses than from non-medical students', for a feeling of certainty is likely to arise from a medical training".

Another test is called a classification probability task, and is intended to indicate the degree to which people can discriminate varying shades of uncertainty. There is a collection of cards, with phrases inscribed on them. The subjects are given the cards and told to sort them into piles such that the cards in a given pile have the same meaning. The idea is a simple one. People who have a fairly discriminating view of an uncertain world ought to be able to create more piles, and will view the phrases rather more widely. The phrases were suggested by Sara Lichtenstein (1966). The main dependent variable is an information measure, (H), and we emphasise the number of piles people create.* We got some very interesting correlation, although none of it "made sense", apart from indicating marked (but complex) cultural and individual differences.

The most complex task is a probability assessment questionnaire (Fig 2). This task requires that people make judgements with probabilities. It is patterned after work done by Peterson. The respondent is presented with dichotomous items, ticks the more correct answer and gives a probability that he is right. The idea is that a well calibrated probability assessor will behave as follows. For all those questions to which (say) he assigns 80% probability, he would be right about 80% of the time. The hit rate, is the proportion of times one is correct. Since it is related to the assessed probability, a perfectly calibrated assessor's hit rate proportions and the probabilities should be in line (for 80% assessment it is enough to be right 80% of the time).

Another test is called a classification probability task, and is intended to indicate the degree to which people can discrinimate varying shades of uncertainty. There is a collection of cards, with phrases inscribed on them. The subjects are given the cards and told to sort them into piles, such that the cards in a given pile have the same meaning.

This idea is a simple one. People who have a fairly discriminating view of an uncertain world, ought to be able to create more piles and will view the phrases rather more widely. The phrases were suggested by Sarah Lichtenstein (1966). The main dependent variable is an information measure, (H), and we emphasise the number of piles people create.

We got some very interesting correlations, although none of it "made sense", apart from indicating marked (but complex) cultural and individual differences.

The most complex task is a probability assessment questionnaire (Fig. 2.). This task requires that people make judgements with probabilities. It is patterned after work done by Peterson. The respondent is presented with dichotomous items, ticks the more correct answer and gives a probability that he is right.

The idea is that a well calibrated probability assessor will behave as follows. For all those questions to which (say) he assigns 80% probability, he would be right about 80% of the time. The <u>hit rate</u>, is the proportion of times one is correct.

Since it is related to the assessed probability, a perfectly calibrated assessor's hit rate proportions and the probabilities should be in line (for 80% assessment it is enough to be right 80% of the time).

A subject's "sureness" or "uncertainty" characteristics are expressed by a relation between the hit rate and the assessments given and if these quantities $\underline{c}$ (for calibration: also, Phillips suggested one interpretation of behaviourally manifest propensity) and $\underline{r}$ (the value as it is subjectively assessed) are logarithmically transformed the $\underline{c}/\underline{r}$ data is linearised. Specifically a two parameter, Log.-Odds, psychological hypothesis is proposed.*

$$\text{Log } \frac{c}{1 - c} = \text{L A} + \text{B Log } \frac{r}{1 - r}$$

and programs have been written to carry out a Bayesian regression analysis on the data from different subjects and to estimate the values of the two parameters (A, B) as the mode of the a posteriori distribution over these quantities. In general, the $\underline{r}$, $\underline{c}$, plots for different values of A and B are unsymmetrical (Fig 3).

---

* <u>Comment</u>: The odds (as in betting) is a ratio of probabilities. If p is the probability of an event then $\frac{p}{1 - p}$ is its odds. The logarithmic form is additive rather than multiplicative.

Some aspects of the characterisations are shown in Fig 4 which represents c/r plots transposed into the A, B, plane. Subjects characterised in the upper right quadrant are "too unsure" (assessed probability is 0.6 even though he is right in 80% of the test trials, for example). The converse ("too sure") applies to a subject characterised in the lower left quadrant.

A subject characterised in the upper left quadrant or characterised in the lower right quadrant of Fig 4 shows an r dependent behaviour. In the upper left quadrant subjects are "too unsure" up to a critical value r^{ϕ} of r, and beyond that point they are "too sure" (counting a high probability as probability of 1). A subject characterised in the lower right quadrant is "too unsure" above $r = r^{\phi}$ and "too sure" below $r = r^{\phi}$. The value of r^{ϕ} is the value of r at the point where the r/c curve intersects the ideally calibrated assessment curve (drawn in all quadrants as a diagonal).

Baker recalled his experience with subjects using his automatically normalised confidence estimation system (for estimates over four alternatives at any trial). For confidence estimates (subjective probabilities) between 0.25 and 0.75 people are, or soon become, excellent assessors but "once they estimate the probability of an alternative (being correct) as 0.75 or 0.8 they opt for probability 1".

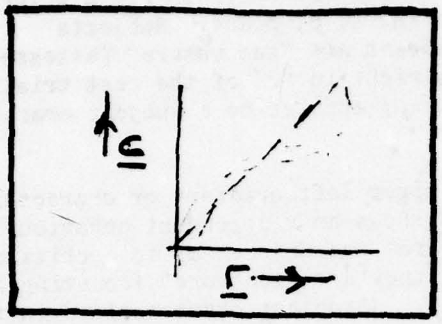
Dr Phillips cited reasons why he could not expect his questionnaire analysis to pick out exactly this effect and there was some discussion of the technical details underlying the questionnaire design which concluded when Tversky asked "In a few words do you think people are over confident in these tasks".

"Beyond r^{ϕ}, yes. The location of r^{ϕ} varies from person to person but for probabilities above r^{ϕ} they are too confident too sure of themselves. r^{ϕ} is usually pretty low ÷ 0.70 or 0.75, for almost everyone. It is nearer 50/50 than it is to 1. That supports Kahneman's observation this morning".

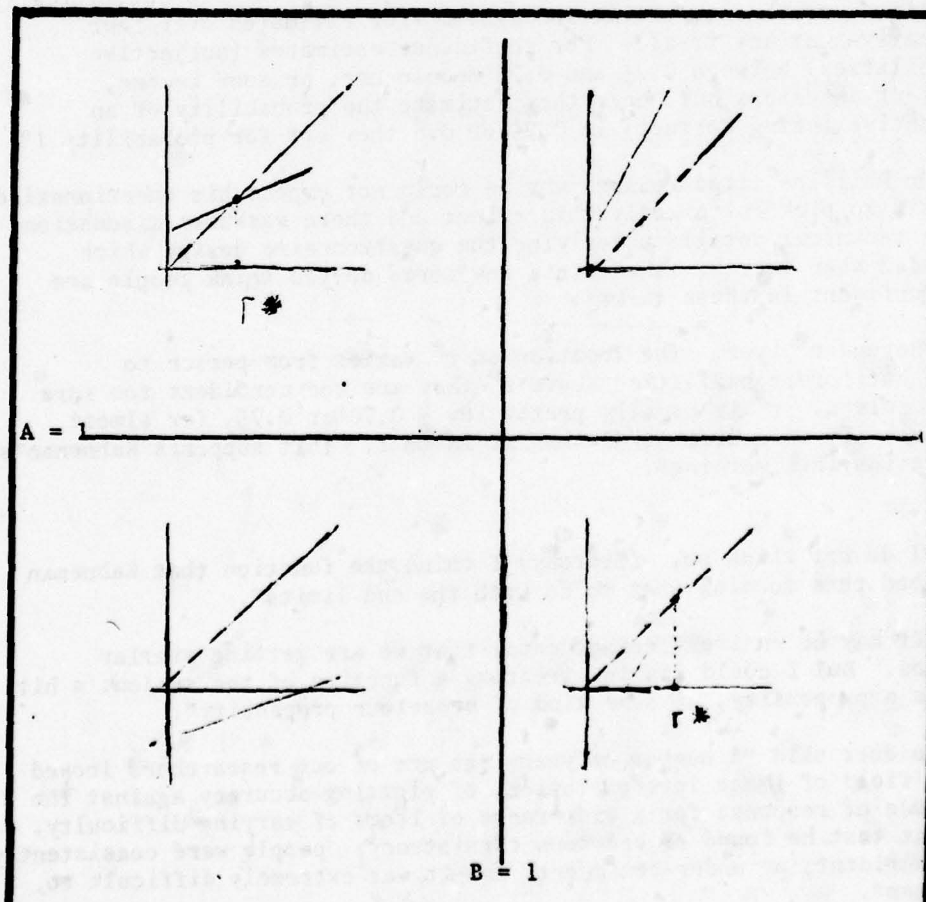
"I do not think so. (Tversky) I think the function that Kahneman described this morning what to do with the end limits".

"It may be entirely coincidental that we are getting similar pictures. But I could imagine treating a function of the subject's hit rate as a propensity, as some kind of behaviour propensity".

Zeidner said "A number of years ago one of our researchers looked at the field of image interpretation, of plotting accuracy against the certitude of response for a wide range of items of varying difficulty. For that test he found an enormous consistency; people were consistently over-confident, or under-confident, and it was extremely difficult to move them".



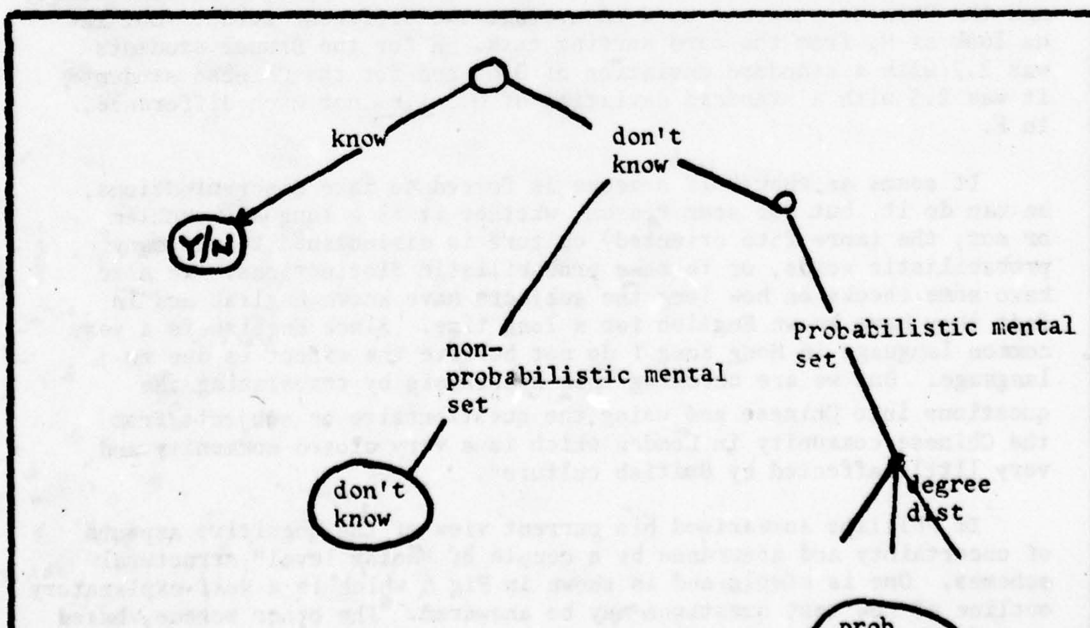
Section 2.11 Fig 3



Section 2. 11. Fig 4

	Brunel Students		Chinese Students	
Y/N x Don't know	-0.49		-0.72	
Prob x Y/N	-0.82		-0.83	
Don't Know x Prob	+0.11		+0.33	
H. x Y/N	-0.43		+0.19	
	Mean	SD	Mean	SD
Y/N	25	9.6	31	10
Hit rate	0.55		0.58	
Don't know	5.9		6.2	
Prob Words	9.2	6	5.6	5
H	2.70	0.62	2.5	0.8

Section 2.11 Fig 5



Section 2.11 Fig 6.

"So far, Lichtenstein and I have analysed the data that we have obtained from about fifty Brunel students and about fifty students in the University of Hong Kong, and I can give the results. Fig 5 shows the correlation between a number of variables. For example: between numbers saying yes/no, or don't know, (for all questions, correct response known and unknown), and the number of probability words, as we might expect, the more likely a person is to use yes/no, the less likely he is to use probabilistic phrases. The correlation between "don't know" and "probability words" is rather interesting because it suggests that there is not a great deal of connection between not knowing and probability words. Considering the classification task; the correlation of the H measure with the tendency to say yes, or no, is -0.43, for Brunel students and +0.19 for the Chinese students, that is a significant negative correlation for Brunel students but virtually no correlation for the Chinese students.

Also, in Fig 5, instead of correlation, I show various means and standard deviations.

Looking at the category "don't know", there is hardly any difference at all, 5.9, and 6.2. If we look at the number of probability words, the Brunel students used about 9.2 with a standard deviation of 6, versus 5.6 with a standard deviation of about 5. The Brunel students therefore, as expected, used more probability words.

I counted the number of different probability words, and it turns out that the Brunel students used on average about seven different words, and the Chinese students used on average two different words. But let us look at H, from the card sorting task. H for the Brunel students was 2.7 with a standard deviation of 0.62 and for the Chinese students it was 2.5 with a standard deviation of 0.8 i.e. not much difference, in H.

It seems as though if someone is forced to make discriminations, he can do it, but for some reason, whether it is a language problem or not, the (more fate oriented) culture is disinclined to use many probabilistic words, or to make probabilistic distinctions. We also have some checks on how long the subjects have known English and in fact they have known English for a long time. Since English is a very common language in Hong Kong I do not believe the effect is due to language. But we are checking this hypothesis by translating the questions into Chinese and using the questionnaire on subjects from the Chinese community in London which is a very closed community and very little affected by British culture".

Dr Phillips summarised his current view of the cognitive aspects of uncertainty and assurance by a couple of "molar level" structural schemes. One is simple and is shown in Fig 6 which is a self-explanatory outline of how test questions may be answered. The other scheme, based on a factor analytic treatment was briefly discussed with the aid of a 3-dimensional model which is too elaborate to reproduce. He proceeded to discuss some further findings, described as "disquieting" yet, for all that, inherently interesting and of potential value.

"Since we compute the H measure on those card-sorting tasks it dawned on me that we should compute the H measure for the last questionnaire. (After all, people can use lots of different probabilities, or very few probabilities, so to treat each probability in a discrete category and to see how often it was used would be an interesting guide to how people make discriminations on a numerical scale). I assumed, of course, that the ability to discriminate on a verbal scale might be correlated with discrimination on a numerical scale. It does not seem to be; there is virtually a zero correlation :

Next (and this was bound to work) look at the hit rate on the known questions for the first task, and the hit rate on the "known" questions on the probability assessment questionnaire. For all those twenty-something known questions, how frequently was the person right? How frequently when a person said 100 percent on the last questionnaire was he right? Zero correlation again!

That suggests, of course, a fourth dimension in the factor analytic scheme which discriminates between numerical and verbal ways of handling uncertainty. That may account for some of the resistance to decision analysis in this country, where there is a very verbal culture, certainly among civil servants. Top people in the civil service tend to come from Oxford or Cambridge, and have studied in arts faculties rather than in science faculties, and they are a bit suspicious of anything numerical. I know of some ministers who refuse to look at graphs, or even numbers and tables.

It must be possible somehow to look at the structure of the language to get some clue about the way an entire culture can think in probabilistic terms. If there are few words available, then that has some influence on thought. One method emerges from back translating questionnaire materials. I took English words indicating uncertainty and dictionary translated them into French. Next, I looked up the French word in the other half of the dictionary, to see if it back-translated into the original word, and if it did, I kept it on my list. I started out with twelve English words, and ended up with twelve French words, but taking those twelve French words and applying the same process (so retaining only those English words that translated into French), I got twice as many English words + 24, which suggests that by double back translating, English-French-English, and then French-English-French the reduction in words that occurs, or the multiplication of words that occurs, might be used as an index of the relative numbers of words that are available in the two languages".

Kahneman said: "There are more in English".

"Yes, but it is not so much numbers as usage".

"I have heard Frenchmen comment that Cartesian thinking is really not very conducive to probabilistic thinking. It is a comment that may relate to the implementation of decision analysis and probabilistic thinking in France. Somebody who is in the business of decision analysis told me that a rationalist Cartesian atmosphere does not lead to probabilistic thinking".

Two other pieces of work are in progress, "investigating the personality and cognitive variables relevant to uncertainty and trying to answer the reasonable question of whether or not the probability assessment questionnaire bears on anything at all".

Tversky commented, "The concept of uncertainty has quite obviously several facets. One is how much you feel you know about what is going on around you; how much you are supposed to know about what can happen. In this respect you can have a very fatalistic orientation as regards uncertainty. For example, to go back to Greek culture. The Gods determine the system which is therefore quite quite deterministic in many ways, except that the Gods are capricious, so we do not know after all. That is one type of uncertainty which could be reflected in judgement.

The other is more pragmatic, and closer to decision analysis or implementation; there is an estimable but irreducible uncertainty. We have to sort out those things of which we know more, and those things of which we know less and classify them, and relate them to the worlds pseudo probabilistic language. It occurs to me that it is possible that those two really go in opposite directions ; back to Phillips original cultural hypothesis. The fatalistic culture may be higher in uncertainty in the first sense, and lower in uncertainty in the second sense, whereas the Anglo-Saxon culture, which is more pragmatic, would be higher in uncertainty because we do not know what is to happen, and the oriental culture would be different again".

"Surely, I have spent time talking to people who know about their cultures and it is fairly clear that the Gods are not capricious in China.

It is well known that the Chinese are inveterate gamblers ; which leads me to reinforce the statement that just because somebody is interested in gampling, and is a keen gambler, that does not mean that he has any appreciation at all of uncertainty."

Kahneman thought "It would be interesting to compare the various games of chance. Horse racing, for example, is mostly blind luck ; yet people look on it as a skill. There are other gambling games which need greater skill as against games of chance, where there is no pretence of skill".

"We started such a study, but there is hardly any reliable literature. It simply raised the fascinating question of what kinds of very fundamental decision-theoretic models are being applied in a culture where the whole view of uncertainty is totally different to the view that we have. As a matter of fact, there is no Chinese word for probability although there is a word for possibility".

2.12. Dr H de Swart (in collaboration with Dr A Dirkswager) "Decision Making and Concept Learning: Remarks and Speculations

The research described by Dr de Swart and Dr Dirkzwager has been carried out in the context of 3 departmental groups. One group consists of psychophysiologists, another is concerned with computer assisted instruction (henceforward, CAI) and the last group works in the field of higher cognitive information processing notably concept learning (henceforward CL). De Swart gave an account of research involving interaction between these different fields.

"A concept learning (CL) task can be seen as a decision situation, and as a result our department became interested in information gathering and integration (especially the Bayesian approach to information processing in judgement). Naturally, also, in the "bookbag and pokerchip" paradign. It is important to note, that our primary aim was to identify the variables that influence judgement not to test the adequacy of Bayes' theorem. Like others, we found that indices of our subjects performance pointed in the direction of the Bayes theoretic prediction but far from optimally. About the same time the "bookbag and pokerchip paradigm" came under criticism. For example.

(a) Shanteau (1970, 1972) tested the adequacy of Anderson's integration model (e.g. 1971) in a situation that resembles the "bookbag and pokerchip" situation, although he agreed (Shanteau, 1975) that the test was not adequate. Shanteau (1975) proposed another method to compare the Anderson and Bayesian model. However, by this method the adequacy of each model to describe decisions is not tested directly.

(b) Others said that subjects react in a "bookbag and pokerchip" situation, only to the relative frequency of the chips and not to the Bayesian variables.

(c) Finally the representativeness of the "bookbag and pokerchip" paradigm was questioned".

2.12.1. Quantitative Frequency

A short and simple pilot experiment was performed in connection with relative frequency.

Subjects were confronted with samples of red and blue chips, written on a card. The samples varied with respect to the relative frequency, the numerical difference between both types of chips and the sample size.* The subjects were asked to order the samples

---

* A "bookbag and pokerchip" experiment is defined as follows : There are M mutually exclusive and exhaustive states of the world or bookbags, $B_1, B_2, \ldots, B_i, \ldots, B_M$ (mostly only B_1 and B_2). Each bag is composed of N different coloured chips $C_1, C_2, \ldots, C_j, \ldots, C_N$ (mostly 2, e.g. red and blue chips). From one of the bags a sample D is drawn (with replacement), e.g. r red and b blue chips. The subject is informed about the probabilities $P(C_j/B_i)$ and $P(B_i)$, as well as about the numbers M, N, r and b. Given this information the subject has to estimate $P(B_i/D)$.

with respect to their usefulness for a decision among symmetrically composed bags. Two conditions for the diagnosticy of the bags were used (p_r = .6; .8) to test the following hypotheses.

(A) That if subjects react according to a Bayes theoretic prediction they must order in accordance with (r - b); if they react to the relative frequency, the ordering has to be in accordance with r/b, and if sample size is the determining factor, the ordering has to be in accordance with this variable.

(B) That in both conditions each subject has to order the samples in the same way.

Hypothesis (B) was confirmed in the case of most subjects but not all of them. With respect to hypothesis (A), the data was very confusing. Each ordering was favoured by some subjects, but others gave mixed orderings. We could not give a satisfactory explanation for this behaviour and concluded that subjects differ in the kind of information they want for their decisions; we were, however, unable to give an explanation for why this was so.

2.12.1. Representativeness

Dr De Swart made the following remarks about the extent to which the "bookbag and pokerchip" paradigm is representative.

"(a) We performed a symmetric binomial "bookbag and pokerchip" experiment, employing instead of two bags, two kinds of people, having more or less favourable than unfavourable qualities. The percentage of each kind was given by instructions. By this set up, the diagnosticity of the alternative states of the world was varied (p_r = .6, .7, .8, .9). The analogues of the red and blue chips were favourable and unfavourable adjectives. The overall result was that subjects' performance approximated the Bayes' theoretic form and less suboptimally than is usual in "bookbag and pokerchip" experiments.*

(b) Recently, we carried out an experiment with a more or less lifelike condition distributions within each condition of the data generator were employed. The results indicated that if subjects are acquainted with the distribution of the data generator then the "real life" variable is of no influence with respect to suboptimality; if the subjects are not acquainted with the distribution of the data generator, then the "real life" variable is influential ie. a less lifelike situation causes more suboptimality (de Swart and Tonkens 1975).

(c) Partly as a result of these studies my colleagues in the CAI group became interested in valid estimates of (subjective) probabilities, in situations where subjects have to choose among alternatives and assign a probability estimate, which reflects their certainty or confidence that the chosen alternative(s) is (are) the correct one(s). They searched for an experimental situation, which satisfied two conditions. First, it pays to report true subjective probabilities, while dishonesty is punished. Second, the situation has to be as life-like as possible. Unless the subject is a bad loser, the risks in a real life situation are

probably quite different from a simulation and the influence of this difference is often mentioned as one of the determinants which results in suboptimality in the Bayesian approach to the study of information processing in judgement (a.o: Slovic and Lichtenstein 1971).

It was assumed that an examination meets these conditions. It seemed reasonable to expect that all students would attempt to gain a high score and that the correctness of alternative answers to a question could be assessed straightforwardly."

2.12.3. Decision Making on the basis of subjective Probabilities

Shuford, Massengill and Albert (1966) have provided a paradigm for scoring psychological and educational tests. According to Shuford, a testee maximises his expected score if and only if he reports his true (subjective) probabilities. Dirkzwager (1975) developed an information theoretic scoring formula which meets Shuford's condition.

"This scoring method is based on the idea that if a testee has perfectly learned the subject matter of the examination then there will be an agreement between the testee and the examiner regarding his probability assignments with respect to the alternative answers. Stated differently, if the subject and the examiner are considered as discussion partners, who agree with each other, then they have the same probability distribution over the cluster of messages (statements or arguments), and the mean information transmitted in each direction equals the uncertainty of each of the participants. However, if the probability distributions of the particilants differ, the testee and the examiner will be surprised by the relative frequencies of each other's messages received during the conversation.

Dirkzwager (1975) derived the following formula which is stated for an examination with 5 response alternative questions.

$$\underline{\text{Items Score}} = \left(\sum_{i=1}^{5} q_i \; {}^{5}\mathrm{lg} \frac{p_i}{q_i} + 1 \right) \times 100 \quad \ldots\ldots (1)$$

where i = the number of the alternative (i = 1, 2, 3, 4, 5)

p_i = S' (subjective) probability for alternative i

q_i = the probability of alternative i as assigned by E

In case there is only one correct alternative, the Items Score reduces to

$$({}^{5}\mathrm{lg}\ p_i + 1) \times 100$$

for the correct alternative. This method was called "Multiple Evaluation" (ME) as opposed to "Multiple Choice" (MC) and it was assumed that ME reduces the testee's gambling propensity. If the testee is not quite sure about the correct alternative, he may get credit even though he is not forced to gamble. If however, his (subjective) probabilities for the alternatives do not differ, then he gets no credit points at all.

The method was applied to an examination called Research Method 1 170 3rd year students in psychology participated. To our astonishment we found a large number of mistaken alternatives judged 100% correct and, consequently, a large number of correct alternatives, assigned a probability of zero. The reason for this was not quite clear, especially as no correlation was found between the tendency to give strong opinions and "knowledge" as measured with the Multiple Choice method, taking the most probable alternative as the one 'chosen'. An example (one of the students, who was often very certain about the mistaken alternatives is shown in Fig 1.

This attitude led to a mean score of -218.5, whereas the score would have been 0.0 if this subject had admitted that he knew hardly anything about the subject matter and, consequently, had assigned probabilities of .02 to each alternative. We assumed that this tendency to assign extreme probabilities was a habit acquired by students who were acquainted with Multiple choice tests, where the best strategy is to pick out the most probable alternative independent of ones knowledge: the answer "I don't know" or "I don't know for sure" is not allowed or even punished. This habit could interfere with the capability to come to the right decision on the probability to be reported in a Multiple Evaluation test. So we organised a second examination for those students who failed the first one and sent them a letter explaining the relation between the two methods and giving examples of the pay offs (item score) as related to the probabilities they would report. The results were now quite successful; this time the multiple evaluation scoring method could be used (only 12 of the 50 students failed their examination this time).

We concluded that students (or anybody for that matter) are able to learn to make the right decision in critical situations, even if this optimal decision implies reporting that one has insufficient information or knowledge of the subject matter to give one decisive expert opinion".

2.12.4. Concept Learning

Dr De Swart described some recent exploratory experiments on concept learning (CL) one of which he regards as relevant to the field of decision making

"We used the classic attribute CL paradigm with one exception: the feedback was not deterministic, but probabilistic. That is to say, the subjects are not reliably informed to which class an instance (stimulus) belongs: which seems to resemble CL in real

QUESTION	ALTERNATIVES					UNCERTAINTY	ME.	MC.
	1	2	3	4	5			
1	0	100	0	0	0	-143.4	100.0	1.0
2	0	100	0	0	0	-143.4	-329.2	0.0
3	0	40	60	0	0	-75.3	43.1	0.0
4	0	0	100	0	0	-143.4	-329.2	0.0
5	100	0	0	0	0	-143.4	-329.2	0.0
6	0	0	100	0	0	-143.4	-329.2	0.0
7	100	0	0	0	0	-143.4	100.0	1.0
8	0	100	0	0	0	-143.4	-329.2	0.0
9	0	100	0	0	0	-143.4	-329.2	0.0
10	0	0	100	0	0	-143.4	-329.2	0.0
11	0	100	0	0	0	-143.4	-329.2	0.0
12	100	0	0	0	0	-143.4	-329.2	0.0
13	0	0	100	0	0	-143.4	100.0	1.0
14	40	0	0	0	60	-75.3	-329.2	0.0
15	0	0	0	0	100	-143.4	-329.2	0.0
						-134.3	-218.5	0.0

Section 2.12.: Fig 1. Example of the performance of a student, who was very certain about the wrong alternatives. For explanation see text.

life situations. The subject's prediction can be considered as a decision. The subjects were informed about the probabilities. The results show that subjects could not learn the concept in this situation (we stopped after 160 trials; in the deterministic situation, they need about 40 trials). In a slightly different situation (providing only probabilistic feedback about the positive instances) the subjects could learn the concept, although they needed more trials than in the classic deterministic situation. A second result was that they changed their strategy. As is well known, subjects learning a conjunctive concept hardly learn from the negative instances. However, in this situation with probabilistic feedback about the positive instances, they extract information from the negative instances".

2.12.5. Speculations

Dr de Swart considers the results, so far, inconclusive, but the following statements can be made.

(a) Subjects differ in the kind of information they need for a decision.

(b) If subjects encounter an experimental situation, with which they are acquainted, or which is structured in a way that makes sense to them then they can perform as required (a decision, learning a concept, etc.)

(c) The problem they encounter influences their solution strategy.

"We favour the idea that we cannot make any progress before we know more about the subject's internal representation of the outside world, and that the recently developed approach used in semantic memory research provides a vehicle for gaining some insight (especially the work of Collins and Loftus (1974)*; and Smith, Shoben and Rips (1974)* who try to establish relationships between attitudes and concepts, among concepts etc. by classification tasks which could be considered as decision situations). Their findings may help us to map internal representations. We speculate that after we know more about the internal representations, we shall gain more insight into :

(A) the kind of information subjects need to make a decision

(B) the way in which the outside world is structured

(C) the way in which the outside world has to be structured to get optimal learning strategies

(D) the way in which the internal representations have to be altered to get, for example, optimal decisions.

I think these speculations are in accordance with the remarks of Prof Kahneman when he proposed to tackle the decision problem by the human engineering approach.

References

Collins, A.M. & E.F. Loftus. "A spreading-activation theory of semantic processing." Internal Publication, 1975.

De Swart, J.H. "Opinion revision and the influence of corresponding and conflicting adjectives." Acta Psychologica, 1971, 35, 198-206

De Swart, J.H. & E. Das-Smaal. "Strategies in ordering samples in a bookbag and pokerchip situation; a preliminary study." Internal publication, 1975.

De Swart, J.H. & R.I.G. Tonkens. "The influence of order of presentation and characteristics of the datagenerator on opinion revision." 1975, in press.

Dirkzwager, A. "Computer-based testing with automatic scoring based on subjective probabilities." Accepted paper for IFIP 2nd World Conference "Computers in Education" , Marseille 1975.

Shanteau, J.C. "An additive model for sequential decision making." Journal of Experimental Psychology, 1970, 85, 181-191.

Shanteau, J.C. "Descriptive versus normative models of sequential decision making." Journal of Experimental Psychology, 1972, 93, 63-68.

Shanteau, J.C. "Averaging versus multiplying combination rules of inference judgment." Acta Psychologica, 1975, 39, 83-89.

Shuford, E.H., H. Massengill & G. Albert. "Admissable probability measurement procedures." Psychometrika, 1966, 31, 125.

Slovic, R. & S. Lichtenstein. "Comparison of Bayesian and regression approaches to the study of information processing in judgment." Organizational behaviour and human performance, 1971, 6, 649-744.

Smith, E.E., E.J. Shoben & L.J. Rips. "Structure and process in semantic memory." Psychological Review, 1974, 81, 214-241.

3. An Overview

After scanning the relevant literature, even the quite recent literature, one comes away with the impression that Decision Theory is greatly concerned with elaborate "multi dimensional beam balancing". The weights in this analogy stand for probability numbers, and the "mechanical advantage" of the levers about a fulcrum stand for utilities. Further, this mathematical apparatus seems, on the same cursory evidence, to be given tidily demarcated interpretations; namely, the normative or prescriptive interpretation recommending what ought to be done and the descriptive interpretation which specifies what is done and thus, if correct, has predictive power. Though obviously a parody, these perspectives are seen as prevalent (and are disputed) by experts in the field (Nigel Howard Paradoxes of Rationality MIT Press 1973, for example), so the impression is not altogether naive. Insofar as the present conference is representative of informed thought in this area, it is clear that neither perspective is seriously entertained. Moreover, the emergent decision theory, more properly, perhaps, decision science, is both more sophisticated and of greater practical value than it might have seemed to be.

In the first place the two neat and tidy interpretations exist only as polarities and in situations concerned with sizeable decision processes they are merged or smudged. Typically, the decision analyst, (Stael von Holstein) the utility assessor (Keeney, Wendt) the guide who points to frequently occuring misperceptions (Tversky) a leader in group dialogue (Hogarth) or the man machine system designer (Baker) starts off with a descriptive theory (which may either be locally applicable and accurate or global and approximate); together with some logical principles, Hume's (and Howard's) "logically necessary" and empirically uninformative statements. By one means or another, discussion or man machine interaction, for example, certain norms are prescribed to fit the expert or decision maker, either a group or an individual. These have roughly the status of a "prescriptive theory" tailored to the situation and individualised. Typically, the agreed compromise is reached in stages, during which specially tailored descriptive hypotheses arise and are used, tentatively, as predictors that guide the dialogue or interaction. Under these circumstances, the descriptive and prescriptive interpretations are effectively inseparable.

Sometimes, of course, a clearcut polarity does exist. If so, the type of mathematical representation parodied as "beam balancing" is not advanced as a normative theory. It is probably true to say that there is no general normative theory of decision (only the specialised prescriptions mooted in the last paragraph). The mathematical apparatus, including the probability calculus as well as the utility schemes, are taken as logical forms which may be held to determine rational action only if special conditions prevail (see, for example, Tversky's paper, on Contingent Probability).

In contrast, there are several descriptive theories or theories-in-the-making. These theories have (and it may be argued they must have) several of the following properties; sometimes they possess all of these properties.

(a) Certain types of cognitive mechanism are commonly manifest in the course of decision making. Amongst these are the classification and rule imaging styles discussed by Broadbent, Tversky's propensities, the quirks and oddities of behaviour uncovered as a result of sequential experiments (Kleiter). All of these mechanisms are rule obedient and may thus be described. They may or may not follow the standard rules (for example, some propensities adhere to directional or causal, rather than symetric modes of inference).

(b) There are marked individual differences in the type of mechanism brought into play when decisions are made. Possibly, the same individual employs different cognitive tricks upon different occasions and (almost certainly) in different kinds of situation. Similarly, there are cultural biasses (Phillips) with respect to decisive behaviour in comparable situations.

(c) The operation of these mechanisms is manifest in terms of gross effects (for instance, Kahneman's "isolation" effect and his "certainty" effect) which can be given a molar description by augmenting the usual type of representation (Kahneman's "uncertainty weight" augmenting "subjective probability" as a descriptor of the "certainty effect", or Phillips account of "critical points", beyond which people are "too sure"). But the molar representation is underpinned by a molecular level mechanism (as in (a)), or a class of them. In this sense, decision theory is psychologically based at a grain or scale of investigation which deals with the cognitive processes behind behaviour as well as behaviour itself.

(d) In general, descriptive theories are made relative to an individual or group in a situation; that is, the description, at either a molar or at a molecular level, is predictive only if the mechanisms implicated are specified, which may entail "on line" enquiry.

(e) Although human beings are involved in any decision (the psychological orientation of (c) is correct), elements other than human beings may legitimately participate, ie generally, decisions are made by systems having mechanical as well as human components coupled into an appropriate organisation. This view is emphasised by Baker (SIMTOS), by Broadbent's observation of managerial gaming and by Hogarth's discussion of groups (for some aspects not obtrusive in the present proceedings, see also Hogarth R.M. "Process Tracing in Clinical Judgement" Behavioural Science Vol 19 No 5 1974 pp 298-315).

Some consequences of an emergent decision science characterised by an interplay of interpretations and having descriptive models with properties (a), (b), (c), (d), (e), are pursued in the following brief essays. So far as possible the essays fill gaps in the proceedings and occasionally extrapolate the argument to suggest fruitful areas of research and development. The main intention, however, is to integrate Dr Zeidner's original statement (Section 2.1.) and the discussion which took place insofar as integration is not already obvious in the text of Section 2.

4. Discussion, Dialogue, and Exteriorising Cognitive Operations

From Section 3, it is more realistic and convenient to draw a distinction between systemic and descriptive theories than between prescriptive and descriptive theories; the norms or prescriptions that arise as a result of analyst/expert dialogue or decision maker/system interaction are compromise solutions arbitrated by recourse to formal or logical schemes. *

The possibilities are usefully summarised in Fig 1 in which rows represent the analyst's or investigator's intention (a bias to forming a decision system or a bias to forming a descriptive theory of decision making) and the columns stand for preferred methods (either dialogue is allowed or there is an experimental situation in which conditions, inputs and so on are controlled quite rigidly). "Dialogue", in this context, includes much of the interaction of SIMTOS but does not include simple feedback, delivered, for example, in Kleiter's experiments. Entries in the table outline the merits and disadvantages of each combination of intention and method.

Since the advantages of not allowing dialogue during the conduct of an experiment are nearly identical with the advantages of impartiality and objectivity obtained by using a standard behavioural experimental design, no special comment is required on this score. It should only be noted that the table is a gross oversimplification. In fact, the investigators opting for the right hand ("on line", no dialogue) column are not under any illusion about a naive "early behaviouristic or physics like" set up. The situation achieved as a result of using questionnaries (Kahneman and Phillips) or confidence estimation consoles (Kleiter and De Swart) is much more complex than that and no one denied it. For example, Broadbent permits discussion during his gaming experiments provided the game rule is not given away, and, in the classification experiments provided a scheme of classification is not imposed upon the subject. This very informative dialogue is a meta dialogue about the experiment and, as such, it does not interfere with the task in hand. The others glean similar information from retrospection after the experiment or from interviews.

---

* That is, "Decision Theory" gives a mathematical interpretation only. This is the status assigned to game theory (a fortiori to decision theory) by Rappaport and tersely summed up in a recent paper (Rappaport. A. "Some comments on Prisoner's Dilema, Metagames and other solutions" Behavioural Science Vol 20, No 3, pp 206-209). The article is a reply to Robinson's critique of Howard's Game formulation, "Metagame Theory", which appears in the same issue and provoked a relevant discussion because Robinson interprets Howard as advocating the metagame as a prescriptive norm in this particular

Theorist's intention		Interaction (dialogue, including appropriate man machine transactions) but excluding rigid performance feedback.	
		Allowed	Prohibited
	<u>System Orientation.</u> Form operating system (like SIMTOS) or formulate model for use.	Proves effective in decision analysis (Von Holstein) and in SIMTOS like systems, since machine/analyst can learn about respondent and respondents can learn (about decision theory or situation) from analyst/machine. As a result, however, the distinction between tactical and strategic decision is abraded if not altogether denied.	Does not work unless satisfactory theory for the decision maker already exists. This is true for very special situations.
	<u>Descriptive Orientation.</u> Formulate or test model for the decision maker	Model may be valid only if Baker's paradigm (a general thinks like the general's staff) is accepted. If not, the data is contaminated by the intrusion of experimenter as a participant. However, without necessary committment to the paradigm, models of this kind can tie together the descriptive theories mentioned in the lower right hand cell of the table.	Individual (Broadbent Kahneman) intergroup (Phillips) and sequential (Kleiter) differences and dependencies call for complex model of the Decision Maker. Elegant mathematically based theories which simplify decision prove inadequate. The complex models, when available, have the status of "pure" or "unadulterated" descriptive theories of a cognitive mechanism underlying behaviour.

<u>Fig 1</u>

The other possibility that active (and by no means always metalinguistic) dialogue should be encouraged, calls for further analysis, especially since (in accord with most of the participants) I have placed certain kinds of mechanical interaction on a par with debate.

Stael von Holstein was the first to state that dialogue, from interviews between a decision analyst and the decision maker or expert, is a more expeditious method of finding out how a decision maker (expert) thinks and reaches decisions than, for example, questionnaire responses. The same point was made, in the context of utility determination, by Keeney and Wendt. Further Hogarth emphasised that group discussion, albeit guided by rules and cautions, is the most potent way of reaching agreement. An acceptable agreement is not simply a consensual aggregation, by vote counting or the like.

In all cases, the dialogue is used to exteriorise normally inaccessible cognitive operations. This certainly informs the analyst or experimenter. If the same data is communicated, and, if necessary, explained to the subject it may also inform him about his own mental activity. As a rule, such a communication is inescapable and is often encouraged; hence, the "adulteration" or "biassing" of the results. Baker and Zeidner pointed out that man/machine as well as man/man interaction may serve the same purpose in SIMTOS, though the interaction in Kleiter's or De Swart's experiments does not do so. The crux of the distinction is probably as follows.

The psychological and systemic breakpoint between the "information gathering phase" and the "combat phase" in SIMTOS (Section 2.2) is not simply marked by the stress of combat action (though this pressure is a significant factor), but, more fundamentally, by the appearance of a genuine man and machine enchange. After this point the respondent uses SIMTOS co-operatively, rather than regarding it as a tool and a store of relevant data. Stress acts as a progenitor of the dialogue.

A few points deserve emphasis since they crucially influence the design of decision systems and possibly also the description of a decision maker.

(a) Objectivity (in the literal sense of regarding the respondent or decision maker as an it-referenced object during an experiment) should not be confused with observability (nor was this confusion apparent; but it is all too easy to fall into the trap).

(b) The observation of cognitive processes underlying behaviour entails exteriorising these processes as verbal exchanges or complex machine oriented behaviours. Direct observation may be mandatory; it looks as though the alternative of inferring cognitive processes from stretches of simple choice behaviour is often an inadequate source of the information needed to build a descriptive theory.

(c) The efficiency of dialogue or machine interaction is well attested; when the intention is systemic (to model a situation or to obtain the meaning attached to value judgements) this expedient is the rule.

(d) Not all machine interaction has a dialogue like quality (by token of which, the breakpoint in a SIMTOS simulation) but all machine interaction can be cast into a dialogue like mode, with increase of available information as the reward.

(e) In the interests of observability, even the descriptive theorist may opt in favour of dialogue (as in (b)) and, if so, an appropriate kind of man/machine interaction has an advantage over man/man interaction. This advantage does not stem from the superficial (and frequently false) impression that a machine is "more impartial" than a man; a sophisticated machine is far from impartial. The important issue is that transactions between a man and a machine are more readily interpretable than verbal transactions, that they can be fully and unambiguously recorded, and these records indicate what biasses exist to be compensated. As above, the "non existence of bias" is usually fictional.

5. Many faces of doubt and preference

This essay, which is more philosophical than the rest, reflects the proceedings but is chiefly intended to pinpoint regions where research and further discussion are desirable.

5.1. Doubt

"What is doubt, or, conversely, what is certainty?" This question will be hedged, just at the moment, by a partial reply to the effect that doubt is a (property of a) state of awareness. Whilst conscious, I cannot be utterly doubtful; as in a void.(I could, of course, be at a loss to say "on which side the dice will fall"). Nor can I be utterly certain; like a robot set in a hyperfatalistic culture.(I could,of course, be certain that the decanter is on the table at this instant). To be "conscious, is also to "entertain doubt" but any attempt to quantify the doubt is contingent upon at least two further statements, namely.

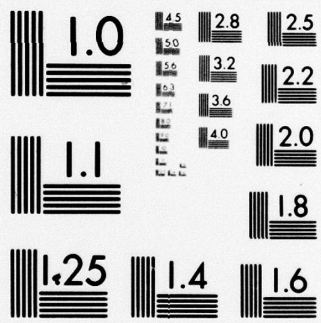
"Who entertains the doubt" (the present candidates are a decision theorist DT and a decision maker DM) and "What is in doubt" (more possibilities are relevant). Having established a framework by answering those questions doubt may be quantified in terms of probabilities (amongst other kinds of measure) and these, in turn, employed to calculate general indices of uncertainty, information and the like. A gross table summarises a few pertinent combinations of statement.

	DT's Doubt	DM's Doubt
About a person's state of mind or belief or information	About DM (his hypotheses and his doubts)	About DT (not discussed but is relevant in a decision system where there are several participants and role of DT is distributed as it is in a decision making team or group)
About an indeterminate world (generally, as in forecasting, about future states)	The DT acts as an oracle or predictor including predictions about the DM's state	Situations in which DM is uncertain of necessity though the DT may not be (for example the DT knows the correct answer to a test question). Also, situations where DM and DT face the same indeterminate world.
About a determinate world (doubt may be researched in principle given sufficient effort)	Either the DT relinquishes his role or treats doubt about DM in this category and, if so, he subscribes to the view that a deterministic descriptive theory of DM behaviour is possible.	Broadbent's situations. A deterministic role is not grasped or DM recollection of past states is imperfect

AD-A077 920 SYSTEM RESEARCH LTD RICHMOND (ENGLAND) F/G 5/1
CURRENT SCIENTIFIC APPROACHES TO DECISION MAKING IN COMPLEX SYS--ETC(U)
APR 76 G PASK DA-ERO-75-G-072
UNCLASSIFIED ARI-TR-76-B1 NL

2 OF 2
ADA
077920

END
DATE
FILMED
1 -80
DDC



MICROCOPY RESOLUTION TEST CHART

Most of the discussion was concerned with circumstances under which the DT is anxious to reduce his doubt about the DM either by formulating and testing a descriptive theory or by estimating the DM's doubt about a determinstic or an indeterministic world. For this purpose, the DM is placed in a measuring situation and there are several categories of measuring situation.

(a) The DM is regarded as an it-referenced object (during the experiment). Indices of hit rate (Phillips) correct response frequency or latency (Kleiter) are used to approximate probabilities and these, in turn, to calculate "objective" uncertainty/information measures. The DMs state of doubt about features of the situation is inferred (together with uncertainty reduction trends,indicative of learning over a series of trials). The inference involves either a standard frequentist interpretation or, in the case of latency, Hick's law (Hick W.E. "The Rate of Gain of Information" Quarterly Journal of Experimental Psychology 1952 Vol 4, pp11, pp26) and Hick's notion of unspecific computation, "Selective work", that reduces the uncertainty engendered by each trial. Either kind of inference presupposes that the DM has a goal to achieve or a problem to solve; perhaps merely alternative-selection. Although the examples cited are germane to the middle right hand cell of the table, recall and recognition latencies may be employed in the context of the lower right hand cell (determinable world).

(b) The DM is regarded subjectively (Stael von Holstein, Phillips) and asked to give reasonable replies (Section 2.11 Fig 1) or subjective probability numbers (Section 2.11 Fig 2) when presented with aleatory questions. Comparable but much more convenient mechanical devices are used by Baker, Kleiter, Stael von Holstein and De Swart, to elicit confidence estimates from the DM. The objective indices are no more than estimators of the DM's state of doubt. As a matter of fact the subjective indices are also simply estimators of DM's state of doubt (the probability numbers do not necessarily reflect DM's true state of belief). But there is a significant training effect: the DM learns, within limits, to use subjective probability assertion as a means for communicating his actual state of doubt and within these limits the DT can say rather than infer that the DM's doubt has changed or remains uniform.

If the DT is a system designer as well as a descriptive theorist the information obtained from these estimates is used to resolve the DT's doubt about an uncertain world (the middle left hand cell of the table) and in turn to help the DM do the same.

It is less commonly recognised (though the point was raised at the conference and met with no opposition) that the middle and lower right hand cells of the table should contain numerous possible entries. These delineate different kinds and species of doubt, which, to some extent, vary independently. Distinction between the kinds and species of doubt is especially important when the decision maker takes part in a man/man or man/machine decision process; a few possibilities are as follows :

(α) Doubt (or certainty) about the problem in hand or the focus of attention. Unless this doubt is nearly zero, at any instant when beliefs of other kinds are estimated, then the estimates in question are of little value. For the most part this kind of doubt was assumed to be resolved at all points of estimation; hence,the

continual emphasis upon common languages and understanding situations. None of this is unreasonable, but it is possible to quantify the extent to which the common language or understood problem exists (which is fortunate if decision making is also intended to adumbrate problem formulation).

(β) Given a goal or problem description which is unambiguous and sufficient to delineate possible outcomes with certainty, doubt about which of these outcomes will occur, or would be correct if selected. This is the doubt/uncertainty information index calculated from Most Confidence estimate responses.

(γ) Given that a problem is solved, doubt about how it is solved, or, retrospectively, doubt about the method used for solving it. Notice, in particular, that this doubt may increase as doubt of type (β) decreases. For example, in a well learned skill, a subjects (β) type doubt may be nearly zero; he is certain about the correct response, or the correct action. But his (γ) type doubt is often very high since he has a vast number of solution methods accumulated by overlearning, any or all of which could be used in solving the problem at hand, and his uncertainty about which method is used (as a better enquiry which group of methods is used) is typically much greater than his uncertainty about the outcome.

These categories of doubt merely exemplify the many kinds of doubt that may be entertained by a subject, and it is appropriate, in this context, to notice the work of Prof Cohen (who was unable to attend the meeting because of a previous commitment) which is overviewed in Behaviour Under Certainty (George Allen and Unwin 1964). The main consequences of the dissection are as follows :

(1) The nature as well as the quantification of uncertainty calls for investigation.

(2) If the DT is a system designer rather than a descriptive theorist then the interpersonal doubts exemplified by the upper right hand cell of the table (DM's doubt about DT beliefs and vice versa) must (and incidentally may) be quantified either to achieve man machine compatibility or to expedite team and group decision processes.

(3) Many kinds of doubt, belief, certainty or whatever can be quantified and there is no reason to believe that doubt is somehow homogeneous. In the context of man/man or man/machine decision systems, a multiplicity of quantification schemes is clearly desirable, and may even be mandatory.

(4) The speciation of doubt is equally important in Broadbent's context, of uncertainty about determinate aspects of decision making. For example, uncertainty of recall is not the same as uncertainty in recognition , though both involve searches through storage indices and contents. Baker and Broadbent discussed some recent work in this field.

(5) The type of belief that is metricised by an index such as a confidence estimate depends upon the question asked; "how" and "why" and "what" questions, for example act differently. As a result, a case can be made for embedding subjective probability/degree of belief elicitation/confidence estimation techniques in a properly formulated logic of commands and questions. There are many candidates; for example, Rescher's (Command) logic, Von Wright's, Harrah's, Belnap's, Aquvist's, Chiarviglio's logics of questioning.

5.2.Utility and Value

A theory of utility and value is "representational" (to borrow a phrase from Luce and Raiffa Games and Decisions Wiley, 1956). Utilities "represent" the worth of some commodity to the decision maker as expressed by his preferences which are obtained beforehand; for example, by comparisons over lotteries (Keeney) or the distribution of points over a taxonomic scheme of contingencies (Wendt). As such, the utilities describe an internal state. But utilities cannot be used to predict preferences; as data about certainties may be used (in a behaviour descriptive theory like Kahneman's) to predict choice and the existence of later states of belief. The utilities are derived from preference statements in the first place, and they are derivable only as a result of obtaining agreement to formal rules, under which the preferred outcome depends upon the decision makers perceived subjective probabilities (albeit uncertainty weighted or whatever). This is not to say, of course, that decision makers do not entertain consistent preferences. It looks as though they often do. During the meeting the basic assumptions and contrivances entailed in eliciting the preferences needed to estimate utility functions was stressed repeatedly (by Keeney, Stael von Holstein, Hogarth and Wendt, in particular) and justified as an expedient for encouraging self awareness on the decision maker's part.

It is also important to ask about the nature of value. This question parallels the opening enquiry of this section. "What is doubt?" and it was raised by Kahneman in the context of utility: "What does it mean for someone to take into account all his assets?" Like doubt, utilities have a complex structure. There are many kinds of value, albeit aggregated for mathematical convenience into a multi-attribute utility function.

Discussion took place outside the framework of formal presentations and, for the most part, outside the meeting chamber. The participants generally conceded the essentially complex structure set out by Hartman (or, in a different vein of explication, by Martin R.M. Intension and Decision Prentice Hall 1963) as justifying the mathematical superstructure of multiattribute and multidimensional utility theory. They also debated the notions more recently advanced by Goguen (1968), "The Logic of Inexact Concepts" Synthese, and the other "Fuzzy Set" theorists (Zadeh, C K Chang, Le Faivre), and the curiously dynamic or action oriented definition of "payoff", which lies at the root of Nigel Howard's (Metagame, option analysis) version of Decision Theory. There was sufficient productive argument to provide grounds for regarding these questions as profitable at other meetings.

6. Personal Styles or Misperceptions and system design

A recurrent theme of the conference is the well attested existence of idiosyncracies and their relation to the relatively stable cognitive mechanisms noted in Section 3(a). Before summarising these facets of behaviour and attempting to set them in the context of decision system design it is worth recalling a further comment (also in Section 3), that descriptive theories of decision behaviour are in vogue (at any rate amongst most of the participants) and are sufficiently sophisticated to have predictive power. It is still possible, however, to adopt several attitudes towards the human being who is described; the following distinction, at any rate, is essential.

(a) If the theory were purely descriptive (like an idealised mechanistic psychological theory) then deviations from a predicted behaviour simply lead to a revision of the theory; the theory embodies the decision theorist's norms; aberrant behaviours, such as over or under estimation of probabilities, are not culpable as failures on the part of the decision maker but (unless due to experimental errors) lead to revision of the theory.

(b) This state of affairs is quite unusual. Generally, the decision maker has agreed to certain norms and prescriptions; the underlining is meant to emphasise that the norms and prescriptions are commonly (Section 3) a locally appropriate compromise; arrived at by discussion or interaction with a decision system and the currently prevailing norms (often open to modification Section 3) are incorporated within the decision theorists descriptive theory or the structure of the decision system. Deviations from the agreed norms are corrected as errors which the human being is sorry to make and, perhaps, unable to remedy (because of limited memory for example, or ingrained overestimation). As such the deviations reflect weaknesses inherent in man. Their appearance may either be compensated, using the current descriptive theory, or it may lead to renegotiation of the norms that are appropriate to this person or group.

(1) Given small span and intelligible sets of exclusive and exhaustive alternatives (as in a limited type of multiple choice situation) human beings who have superficially agreed to act "rationally" as probability estimators are prone to various aberrations. (a) To answer dichotomously or by over or under-estimating or in a manner that fails to reflect their actual degree of belief. (b) If this tendency is avoided, in conditions such as those discussed by Phillips, asserted probabilities vary in a complex manner with "Hit Rate" or "veridicial" probabilities (c) The tendency of (a) can be avoided by providing a game like situation in which much or all of the normalising work is done, almost instantaneously, by an external device; which sums the probability numbers to unity; Baker pointed out that children used his histogram display (which automatically normalises the confidence estimate and displays a correct belief score) as a game to which they become aclimatised; further they gave consistent estimates apart from an overestimation effect ($p > 0.8$ taken as unity) (d) Similar comments apply to the confidence estimation systems, reported by Dirkswager and De Swart (here, the probabilistic assertions are held invariant, the probability weight distributed over the remaining

alternatives is automatically normalised) (e) Naive adults (for example, respondents in motivational research) act in the same way, especially if the correct belief score is of the Shuford type (f) Apart from the provision of external aids to otherwise onerous computation, and restriction of the alternatives, a significant training effect is involved in securing reliable estimates; people learn to use the system but are often able to do so over a short interval of training. (Baker, Zeidner). (g) Although the tendency noted in (a) may thus be minimised or even eliminated, the biasses to over and underestimation persist, even in the simplest situations (h) It seems likely that a descriptive theory (Kahneman, Phillips, Tversky, De Swart, Dirkswager) could account for these biasses.

(2) If the choice situation is elaborated to a lottery involving values attached to options and taking account of these "values" relative to the respondents absolute degree of "wealth" then the situation is much more complicated. Under these circumstances (lotteries as a paradigm) a theory such as Kahneman's and Tversky's is required to account for the estimation behaviour.

(3) Under guessing game and sequential conditions (exemplified by Kleiter's paper) there may be local consistency of choice behaviour but there are numerous violations of classical precepts and subjects do not conform to a simple (sequential decision game) criterion of optimal choice or bet placement. It thus appears that sophisticated extrapolations from simple behaviour are inadequate predictors of actual sequences of simple behaviour, especially if between-person and between-task differences are considered in the analysis of data.

(4) There is ample evidence for the existence of individually distinct and culturally distinguishable modes of decision making (Broadbent, Kleiter, Baker, Zeidner, Stael von Holstein, and Phillips amongst others, though Broadbent and Phillips stress the matter for specific discussion).

(5) In this respect, the position adopted by Tversky and agreed by Kahneman is quite crucial. They show that individual differences and oddities in decision making exist, and that, for a wide and important spectrum of situations, the basically acausal notions of probability are positively counterintuitive and may be so counterintuitive as to be (in an operational sense) <u>unthinkable</u>.

(6) For sure, there are certain coherent calculating rules that are accepted by bodies of decision makers and those influenced by their decisions; there are standard situations in which the concensus over rational behaviour is so complete that it is often mistaken for a rule of nature (rational moves in a 2 player game with a saddle point; the rationality of saying an unbiassed and "imaginary" or "combinatoral" dice is equally likely to fall on any face). Over a much larger class of situations, rational inference is accepted as a useful part of a common language. But rationality, in this sense, is no more nor less than the agreed logical coherence and consistency of computationally provident formal rules which may, or, as in the last paragraph, sometimes may not, be the standard rules.

(7) It is generally acknowledged that small span and intelligible choice sets are mandatory. People are not able to estimate degrees of belief in, or subjective probabilities of, complex contingencies. On the other hand, it can be argued that reliable belief and probability estimates may be obtained by standard methods (improved or corrected by the considerations of (1), (2), (3), (4), (5) above), if a decision situation is broken down into small components and Bayes Rule or its derivations are used to compound and aggregate the results.

This paradigm has a great deal to recommend it providing the Decision maker accepts (or is trained to accept) the prescriptive norms and also accepts (or is trained to accept) the breakdown of the decision situation (for example, in terms of a contingent choice tree). But there is a real danger, chiefly voiced outside the meeting room, that one acceptance may be taken falsely, to imply the other, i.e. that a decision maker accepts the norms appropriate to miniscule decisions, accepts an organisation or machine compounding his estimates, but utterly fails to appreciate the structure of the situation upon which the composition is based; moreover, that he would not accept this structure, if he did understand it.

The point is important because exactly this expedient is often used (explicitly or not) for quanityfing prior opinion, concerning parts of a situation; posterior estimates being fed back by an organisation or a machine (much the same comments are true of other kinds of opinion samples; for example, regarding likelihood ratios). There is not the slightest doubt that arrangements such as this have great practicality (for instance, they lie at the roots of Decision Analysis, military systems of various kinds and have figured in laboratory systems, since the pioneering work of Ward Edwards). Nor is there any doubt that the decomposition expedient is often entirely legitimate (for example, when, in the sense of Stael von Holstein's paper, the situation is decomposable and the Decision Maker has a fair inkling of its structure). Nevertheless a cautionary note is in order. Not all decision makers able to appreciate the norms of local choice are in a position to appreciate the structure of a valid decomposition (here, in effect, we face Hogarth's problems of combining the decisions from several heads).

Under these circumstances, (and with particular reference to (6) and (7) above and to Section 5), it is wise to examine a different, but perfectly respectable, perspective, namely, that "Decision" is not just a matter of "Choice" or expressing "degrees of belief" or "subjective probabilities" or "expectations". On the contrary,"decision" is rightly seen as a subprocesses inextricably bound up with the process of conceiving hypotheses and formulating problem situations (one of Zeidner's points in Section 2.1). The request to make a choice between exhaustive alternatives, for example, introduces an odd type of discontinuity in a flux of mental operations and may be more or less arbitrary (even when the alternatives are presented by the environment). It would be merely perverse to introduce this complication in those cases, already reviewed, where human activity can be neatly partitioned into the prior formulation of a model, its decomposition into choice sets, the valuation of outcomes, and the elicitation of confidence estimates. But it is conjectured that, for many of the complex processes we refer to as "decision", this partitioning is impracticable. If so, the conundrum of Section 2.5, "where is the decision maker" in a "decision system", must be faced fairly and squarely (as it was, for example, by McCulloch and

motivated his interest in redundancy of Potential Command * as not only a biological but also a psychological phenomenon).

Organisations having "redundancy of potential command" (especially teams and groups of people or mechanisms) are little investigated and the topic might be a useful focal point for further meetings, in concert with social psychologists.

(8) Given these precautions it appears practicable to construct either man/man or man/machine decision systems and to guarantee the coherence truth (in the sense of Rescher N. The Coherence Theory of Truth Oxford University Press 1973) of the agreements reached by the decision making components. Further, there will be certain consistent logical or mathematical propositions at the base of any such agreement. Obviously, this is no more than a significant step towards success, since logical truths and coherence based agreements do not necessarily relate to empirical truths or facts about the environment being manipulated. However, the significance of this step can be appraised in the light of the following qualifications.

(a) Depending upon the definition of "uncertain world" which is appropriate to a situation, empirical or correspondence truth (even in the probabilistic sense) may be, in principle, unachievable. For example, if structural indeterminancy (disallowing decomposition) or some kinds of asynchronicity and non stationarity are properly countenanced. then the best that can be done is to replace mere voting consensus, or its analogue in machine matching of human postulates, by a coherence based agreement between human decision makers, often operating with or interacting through a computing machine.

(b) When sufficient features of the environment are predictable (for example, in Baker's SIMTOS or the policy studies of Stael von Holstein) this type of design is acceptable as an improvement upon current practice which, in different ways, meets the objections voiced by Hogarth and Zeidner.

(c) Broadbent's observations highlight the very large extent to which the uncertainty of human beings is an uncertainty about completely determinable quantities. All such uncertainties are due to deficiencies which become remediable under this kind of decision system design, whether they arise from the misconception of deterministic rules (for example, the relation between future income and future expenditure) or the imperfections of human memory (as, for example, in keeping track of the previous values of a large number of different state variables).

Notably, all of the problems raised by Strub and Levit in a recent paper (Strub, M.H., and Levit, R.A. (1974) "Computer Compatibility with Decision Style: Eddies in a Bit Stream". Proc. Human Factors Society 18th Annual Meeting, October 15-17, 1974. Huntsvielle, Alabama, pp 46-49) are tackled by the proposed approach and possible solutions are offered.

---

* Enunciated in McCulloch W.S. Embodiments of Mind MIT Press 1965; detailed in several papers of which the most readily accessible is Kilmer, McCulloch, Brun, Craighill, Peterson "A Cybernetic theory for the Reticular Formation" in Cybernetic Problems in Bionics Ed. Oestreichier, and Moore, Gordon and Breach, 1968 (USAF Bionics Conference Dayton, Ohio 1966).

7. Training and the Design of complex decision systems

Learning is an invariable and unavoidable concommitment of decision making (and probably of any other intellectual activity as well). Its effects may be minimised but, even in serial but independent-trial guessing-experiments, the influence of learning is obtrusive as trend matching, the adoption of a rational strategy, or as superstitious behaviour.

Under more orderly conditions, the respondent learns how to use a measuring technique (Baker, De Swart, in the context of confidence estimation); given more complicated experimental constraints (as in managerial gaming (Broadbent), SIMTOS gaming (Baker) and group decision (Hogarth) the learning is strategic or interpersonal. Stael von Holstein, Keeney and Wendt emphasised the significance of learning to cope with techniques and problems; for example, in the course of "utility assessment". Under all these circumstances, the learner (alias decision maker or expert) is being trained by the situation and the situation has a training function (Zeidner) which may or may not be deliberately engineered.

The practical issue is not, therefore, whether decision makers will be trained "on line" (they are trained willy nilly); it is, instead the issue of whether this kind of training is expeditious. Tversky felt, for example, that a mix of training and coding could lead to overload, misconceptions and interference. He advocated a thorough-going educational effort (in schools or the like) to acquaint decision makers with the tools of decision theory (for example, the calculus of probabilities, their own ingrained misperceptions and limits upon power to estimate) before entering into discussion with a decision analyst or participating in a man/man or man/machine decision system as a decision maker. Broadbent noted that "participatory learning" is not always effective and that it is apt to be superficial (subjects may come to manipulate a situation in ignorance of the underlying rules and principles).

All this bears upon the philosophy of man/machine system design. A prerequisite for securing man/machine compatibility is to match human oddities and idiosyncracies (probably,individual by individual) to machine characteristics.

At one extreme it would be possible to employ a "filtering method" i.e. to design the interface so that it compensates for the human idiosyncracies of an individual user (discovered by pretesting), and to make the machine translate its output and react in the user's terms. Although education is unlikely to homogenise decision makers (and this would hardly be desirable in any case) it presumably is possible to eliminate the behaviours which result in gross contradiction from their repertoire, and to acquaint them with the language of decision theory. Intensive education could thus ease the design of a system based upon the filtering method, and enhance its operational efficiency.

At the other extreme, it is possible to build a training function into a partially mechanised system i.e. to implement "training on the job". This expedient is known to be effective in some cases,but the

proceedings make it clear that this "participating learning" is not always effective and may even be counterproductive due to overload and interference. There are thus arguments to advance in favour of both the "education and filtering" design method and the "on the job" design method.

The views expressed by the advocates of "on line" training and "education" are complementary, rather than at odds. Surely, an educational effort is worthwhile. Surely, also, it is possible to design an on line training system that avoids the pitfalls of misbegotten (especially superficial varieties of) "participation learning"; for example, by demanding explanations for classes of decisive action. One compromise, which is particularly attractive and deserves further discussion and investigation, is a decision system in which the mechanical organisation is able to learn about the learning processes of the human decision makers who interact with it; all the better if they arrive with prior knowledge of how the mechanism performs its computations. In fact, there are grounds for arguing that just such a mutually responsive organisation is a further prerequisite for real man/machine compatibility. It is also true that a system able to sensibly accommodate interpersonal uncertainty indices (mooted as another requirement for man/machine compatibility, in Section 5) must have a basic structure of this kind.

8. Decision Tools, Decision Theory and Decision Systems

Decision theory comprises a collection of mathematical and logical tools. If these abstract devices are given a purely mathematical interpretation they are formally consistent. More than that is required of a general decision theory; a theory must (as Tversky pointed out in particular) have a model (in the model theoretic sense of a representation) in one or more universe of interpretation. This requirement, in turn, calls for establishing measurement and identification procedures that are themselves semantically compatible. Although it has been possible to construct locally interpretable descriptive theories, there does not appear to be any general theory of decision making.

The tools, themselves, are unquestionably useful. Apart from the specific formal tools mentioned, explicitly, by Hogarth there are others, for example, the calculus of probabilities, of utilities, and (as theories proper rather than formal devices) the locally applicable descriptive theories. Hence, taking the stance of a system designer, it seems prudent to set aside the lack of a general theory and the slightly embarrassing question of what its universe of interpretation would be, and to concentrate upon the deployment of the tools that exist.

From this perspective, it is clear that many of the calculations and storage operations required in order to use the decision tools are profitably delegated to computing machinery. (1) The calculations are unduly laborious when performed by human beings (2) Humans are subject to limitations upon data storage, etc., which do not beset machinery (3) Humans are liable to error in retrieval as well as calculation (4) There are systematic human errors in reaching judgements which, according to some of the decision tools, should be unbiassed (5) Humans have ingrained misperceptions (for example, they give causal interpretations, of conditional dependencies) which contradict the formal precepts underlying some of the decision tools (6) Humans have strategic concepts at odds with those of the formal edicts. For these and other reasons it is possible to make a case for mechanising most of the decision tools, and doing so at all levels of sophistication.

Suppose the mechanisation is carried out. If so, then a decision maker is a human being or a group (team) of human beings, with the following characteristics: (A) The Decision maker is given, or has formulated, a Decision Problem (B) The decision maker has agreed to use the collection of decision tools in the context of this problem. (C) The Decision Maker selects, at his own discretion, between the decision tools to be employed for solving different parts of the problem. Having made a selection, the tool concerned is used automatically, and (from (B)) the decision maker is bound to accept the result computed (an action, a likelihood ratio, or whatever) and not to distort or bias the result (which will probably be input to another mechanised decision tool).

Could a real life system be as simple as that? It seems unlikely, for several problems are obtrusive; but, if it were, the cognition of the decision maker would be exteriorised by his selections amongst the menu of decision tools at his disposal.

In fact, the simple configuration is a gross idealisation of reality (a) The behaviours referred to as "selections" are properly stated as questions, enquiries, instructions, etc., the totality of which is a series of man/machine or man/man transactions (b) Questioning, enquiry, etc., is not confined to the Decision Maker. If the computing machine is to operate effectively, it must demand information, valuation, etc. (c) The act of procuring a decision is usually a series of transactions in which the results obtained by executing one of the decision tools indicates the selection of a further decision tool (d) In order to operate at all, the machine must be informed of the problem specification and may, in turn, modify it, so that decision tools can be applied (e) Under these circumstances, the decision process is legitimately said to be distributed between the Decision Maker and the machinery, rather than isolated in the Decision Maker who used the machinery as a slave. In particular, the mutual learning design of Section 7, can be realistically assumed to exist.

These apparent complications actually lead to a fascinating possibility. Although it may be impracticable, in the current state of the art, to construct a complete and general theory of human decision making there are good grounds for believing that a general decision theory of man/machine mutualism is not so far distant. The trick is that something reasonably called a decision process is exteriorised as a behaviour, by means of the transactions between the Decision Maker and the machine. It is entirely plausible to imagine a Decision Theory of such behaviours. True, it is not a Decision Theory of human beings, since the machine is not (as in (e)) just a slave but a co-operative partner in decision. But it <u>is</u> a general <u>decision theory</u>, at least for the exteriorised behaviours manifest in the hybrid system. Its <u>universe of interpretation</u> is precisely the set of transactions between the Decision Maker and the machine that realise decision tools.*

This, alone, appears to justify systems such as SIMTOS, not only as devices with pragmatic value, but also as the experimental milieu required in order to test the postulates of such a theory. Within this environment tactics and strategy are inextricably mingled; the former is required if only to exteriorise transactions as behaviours; the latter, reflects the dogma that the machinery is a mutualist rather than a moron. The system is an instrument for decision; the existence of such systems opens up possibilities which seem for me at least, (and it is probably fair to say, for most of the participants) both exciting and of immense potential value.

---

* This point of view is developed in the Decision Maker's aspect of Conversation Theory (Pask, G. <u>Conversation Cognition and Learning</u> Elsevier Press 1975 and in <u>Conversation Theory: Applications in Education and Epistemology</u>, Elsevier Press 1976).